

Symmetric Properties and Subspace Degradations of Linear Operator Channels over Finite Fields

Shenghao Yang, Siu-Wai Ho, Jin Meng, and En-hui Yang, *Fellow*, *IEEE*

Abstract

Motivated by the communication through a network employing linear network coding, linear operator channels (LOCs) over finite fields are studied with arbitrarily distributed transfer matrices. Some intrinsic symmetric properties of LOCs are revealed and are used to simplify transition matrix computation and input distribution optimization. Subspace coding for LOCs is studied with the help of the symmetric properties. Our results demonstrate that using constant-dimensional subspace coding are good enough for many typical parameters. For LOCs satisfying certain constraints, the optimal subspace coding is constant-dimensional. Simple method is derived to find an optimal constant-dimensional input distribution, as well as the maximum achievable rate using constant-dimensional subspace coding.

Index Terms

linear operator channel, linear network coding, subspace coding

I. INTRODUCTION

A linear operator channel (LOC), also called multiplicative matrix channel, with input X and output Y is given by

$$Y = XH,$$

where $X \in \mathbb{F}^{T \times M}$, $Y \in \mathbb{F}^{T \times N}$ and $H \in \mathbb{F}^{M \times N}$ are matrices over a finite field \mathbb{F} with q elements. H is called the transfer matrix of the channel. We assume the noncoherent transmission that the instances of H are not a priori for either the transmitter or the receiver.

A LOC models the communication through a network employing linear network coding [1], [2]. Koetter and Kschischang [3] observed that the vector space spanned by the column vectors in Y is always a subspace of

S. Yang is with Institute of Network Coding, The Chinese University of Hong Kong. (e-mail: shyang@inc.cuhk.edu.hk)

S.-W. Ho is with Institute for Telecommunications Research, University of South Australia, Australia. (e-mail: siuwai.ho@unisa.edu.au)

J. Meng and E.-h. Yang are with Department of Electrical and Computer Engineering, Waterloo University, Waterloo, ON, Canada. (e-mails: {j4meng, ehyang}@uwaterloo.ca)

the vector space spanned by the column vectors in X , and proposed to modulate information by subspaces for communication through LOCs, which is called subspace coding. They explored the subspace coding problem for one use of LOCs [3]. Thereafter, subspace coding generated a lot of research interests, but there are still some fundamental things about subspaces coding and LOCs unclear. For example, what is the relation between LOCs and subspace channels, and when is subspace coding optimal for LOCs? Generally, we may want to know how to design subspace codes for multiple uses of the channel.

Towards better understanding the coding problem for linear network coding, a systematic study of LOCs becomes necessary. Existing works study special distributions of H . When $M = N$, Silva *et al.* [4] studied that H are uniformly chosen from all full rank $M \times M$ matrices. Jafari *et al.* [5] studied that H contains uniformly i.i.d. components. However, in typical network coding applications, the transfer matrix H is jointly determined by the dynamics of the network topology, the packet dropping pattern, the randomness in linear network coding [6] and other random factors in the network transmission. So, H can have rank deficiency and correlated components, and the distribution of H is hard to be determined. Though studying a specific distribution could get deeper results, it is arguable how these results can be applied to other distributions of H .

In this paper we focus on the general properties of a LOC with an arbitrary distribution of H . It is well known that if T is much larger than M , parts of X can be used to transmit a known matrix such that the receiver can recover the instances of H . Such a scheme has been widely used for random linear network coding [6] and is asymptotically optimal when T goes to infinity. Here we are interested in the case with mild T , where noncoherent transmission becomes meaningful. Our results, unless otherwise specified, work for general parameters T , M , N and q .

In the first part of this paper, we investigate the symmetric properties of LOCs which hold for any distribution of H . The input and output of a LOC are matrices and have higher dimensions of freedom than those of channels like BSC and BEC. For example, given the distribution of H , directly computing the transition matrix of a LOC has a complexity $\mathcal{O}(q^{TM+MN})$. We show that using some intrinsic symmetric properties of LOCs, we can reduce the complexity of computing the transition matrix to $\mathcal{O}(\min\{M, T\}q^{(\frac{M+N}{2})^2})$.

Moreover, optimizing the mutual information is usually not easy since the number of probability masses exponentially increases with the dimensions of input matrices. An input distribution is said to be α -type if two input matrices spanning the same vector space by their rows are equiprobable. We show that there exists an α -type input that achieves the capacity of any LOCs. An α -type distribution is equivalent to a distribution over the set of subspaces of \mathbb{F}^M with dimension less than or equal to $\min\{M, T\}$. Thus to find an optimal α -type distribution has much less number of variables to fix than to find an optimal distribution over $\mathbb{F}^{T \times M}$.

In the second part of this paper, we study subspace coding for LOCs with the help of the symmetric properties. A subspace degradation of a LOC is defined as the subspace channel induced by the LOC with a given transition probability from subspaces to matrices. We say that a LOC is uniform if the transition probability of the channel only depends on the subspaces spanned by the column vectors of the input and output matrices. We show that a uniform LOC has a unique subspace degradation and the subspace degradation can achieve the same rate as the

LOC.

Our results demonstrate that using constant-dimensional subspace coding may suffice for many scenarios. We show that the gap between the maximum achievable rate of constant-dimensional subspace coding and the maximum achievable rate of subspace coding is at most the maximum information can be transmitted using input and output ranks, which is less than $\log_2 \min\{T, M, N\}$ bit per use. This gap is marginal for typical parameters and diminishing when either q or T goes to infinity. We also show that the optimal subspace coding is constant-dimensional for LOCs with i) the rank of H has positive probability to be any integer from 0 to M and ii) sufficiently large T .

We derive a linear programming to find an optimal constant-dimensional input distribution, as well as the maximum achievable rate using constant-dimensional inputs. In the general case the complexity of this linear programming is linear with the number of subspaces in \mathbb{F}^M with dimension less than or equal to $\min\{M, T\}$. When T is sufficiently large, the optimal dimension is at least the largest rank of H with nonzero probability. For uniform LOCs, the complexity can be reduced to $\min\{T, M\}$.

Parts of the results of this paper have appeared in our conference paper [7] and online in [8]. A random matrix is said to be uniform (for a given rank) if the instances of a same rank are equiprobable. Recently, Nóbrega *et al.* [9] studied LOCs with uniform transfer matrices. A LOC with a uniform transfer matrix is always a uniform LOC defined in this paper, but a uniform LOC may have a non-uniform transfer matrix. We will make it clear by an example in this paper.

The rest of this paper is organized as follows. After introducing some notations, we formally define a LOC in Section III. In Section IV, we introduce some symmetric properties and show how these symmetric properties simplify the study of LOCs. Subspace degradations of a LOC are introduced in Section V, where we also discuss uniform LOCs. The mutual information decomposition of subspace degradations in Section VI is an important result of the symmetric properties. Based on this decomposition, we obtain the results about constant-dimensional subspace coding in Section VII. Uniform LOCs are further discussed in Section VIII.

II. PRELIMINARIES

Let \mathbb{F} be the finite field with q elements, \mathbb{F}^t be the t -dimensional vector space over \mathbb{F} , and $\mathbb{F}^{t \times m}$ be the set of all $t \times m$ matrices over \mathbb{F} . For a matrix \mathbf{X} , let $\text{rk}(\mathbf{X})$ be its rank, let \mathbf{X}^\top be its transpose, and let $\langle \mathbf{X} \rangle$ be its column space, the subspace spanned by the column vectors of \mathbf{X} . Similarly, the row space of \mathbf{X} is denoted by $\langle \mathbf{X}^\top \rangle$. If V is a subspace of U , we write $V \leq U$.

For a discrete random variable X , we use p_X to denote its probability mass function (PMF). For random variable X and Y defined on discrete alphabets \mathcal{X} and \mathcal{Y} , respectively, we write a transition probability (matrix) from \mathcal{X} to \mathcal{Y} as $P_{Y|X}(\mathbf{X}|\mathbf{Y})$, $\mathbf{X} \in \mathcal{X}$ and $\mathbf{Y} \in \mathcal{Y}$. When the context is clear, we may omit the subscript of p_X and $P_{Y|X}$ to simplify the notations.

III. LINEAR OPERATOR CHANNELS

Let T , M and N be positive integers. A LOC with input $X \in \mathbb{F}^{T \times M}$ and output $Y \in \mathbb{F}^{T \times N}$ is given by

$$Y = XH, \quad (1)$$

where H , namely the transfer matrix, is a random matrix distributed over $\mathbb{F}^{M \times N}$. Such a LOC is denoted by $\text{LOC}(H, T)$. For one use of $\text{LOC}(H, T)$, we mean the channel transmits one $T \times M$ matrix.

A communication network employing linear network coding can be modeled by a LOC. The source node encodes its message into batches (also called generations, classes or chunks), each of which containing M packets of T symbols [10], [11]. Network nodes perform linear network coding among the symbols in the same position of the packages in one batch, and the coding coefficients for all the positions are the same. This packetized transmission matches our assumption that the transformation matrix keeps constant for T positions of the packets.

In a general situation, the number of packets received of batch is also a random variable. Here we use a fixed number N of column of Y because i) receiving unbounded number of packets for a batch does not make sense in practice, so we can put a bound N for the number of packets that can be received for a batch; ii) when the number of received packets is smaller than N , we can always make the number of received packets to be N by padding all-zero columns into Y .

We assume that H and X are independent. Under this assumption, the transition probability $P_{Y|X}(\mathbf{Y}|\mathbf{X})$ is given by

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{\mathbf{X}H = \mathbf{Y}\}. \quad (2)$$

A LOC is a *discrete memoryless channel* (DMC). The *capacity* of $\text{LOC}(H, T)$ is

$$C(H, T) = \max_{p_X} I(X; Y).$$

Achieving the capacity generally involves multiple uses of the channel. A block code for $\text{LOC}(H, T)$ is a subset of $(\mathbb{F}^{T \times M})^n$, the n th Cartesian power of $\mathbb{F}^{T \times M}$. Here n is the *length* of the block code. Since the components of codewords are matrices, such a code is called a *matrix code*. The channel capacity of a LOC can be approached by a sequence of matrix codes with $n \rightarrow \infty$.

A. Markov Chains

Let X be a random variable over $\mathbb{F}^{t \times m}$. Let $\text{Pj}(\mathbb{F}^t)$ be the collection of all subspaces of \mathbb{F}^t . Let $\langle X \rangle$ be a random variable over $\text{Pj}(\mathbb{F}^t)$ with

$$p_{\langle X \rangle}(U) = \Pr\{\langle X \rangle = U\} = \sum_{\mathbf{X} \in \mathbb{F}^{t \times m} : \langle \mathbf{X} \rangle = U} p_X(\mathbf{X}). \quad (3)$$

Denote X^\top as a random variable over $\mathbb{F}^{m \times t}$ with $p_{X^\top}(\mathbf{X}^\top) = p_X(\mathbf{X})$. Combining the above notations, $\langle X^\top \rangle$ is a random variable over $\text{Pj}(\mathbb{F}^m)$ with

$$p_{\langle X^\top \rangle}(V) = \sum_{\mathbf{X} \in \mathbb{F}^{t \times m} : \langle \mathbf{X}^\top \rangle = V} p_X(\mathbf{X}).$$

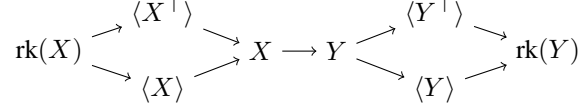


Fig. 1. Random variables and Markov chains related to $\text{LOC}(H, T)$.

Furthermore, denote $\text{rk}(X)$ as a random variable with

$$p_{\text{rk}(X)}(r) = \sum_{\mathbf{X}: \text{rk}(\mathbf{X})=r} p_X(\mathbf{X}). \quad (4)$$

It is easy to see that $\text{rk}(X)$ is a deterministic function of $\langle X \rangle$ ($\langle X^\top \rangle$), and $\langle X \rangle$ ($\langle X^\top \rangle$) is a deterministic function of X .

Now we consider $\text{LOC}(H, T)$ where H has dimension $M \times N$. Applying the above definitions on the input X and the output Y , we obtain the relation between random variables shown in Fig. 1. These random variables are given as the nodes of a directed graph. All the random variables in a directed path form a Markov chain. For example, $\text{rk}(X) \rightarrow \langle X \rangle \rightarrow X \rightarrow Y \rightarrow \langle Y \rangle \rightarrow \text{rk}(Y)$ forms a Markov chain. Let $r, U, \mathbf{X}, \mathbf{Y}, V$ and s be the instances of $\text{rk}(X), \langle X \rangle, X, Y, \langle Y \rangle$ and $\text{rk}(Y)$, respectively. To verify this Markov chain, we only need to check the deterministic relations between these random variables:

$$p(r, U, \mathbf{X}, \mathbf{Y}, V, s) = \begin{cases} p(\mathbf{X}, \mathbf{Y}) & \text{if } \langle \mathbf{X} \rangle = U, \dim(U) = r, \\ & \langle \mathbf{Y} \rangle = V, \dim(V) = s, \\ 0 & \text{o.w.,} \end{cases}$$

$$p_{\text{rk}(X)\langle X \rangle}(r, U) = \begin{cases} p_{\langle X \rangle}(U) & \text{if } \dim(U) = r, \\ 0 & \text{o.w.,} \end{cases}$$

and

$$p_{\langle Y \rangle \text{rk}(Y)}(V, s) = \begin{cases} p_{\langle Y \rangle}(V) & \text{if } \dim(V) = s, \\ 0 & \text{o.w.} \end{cases}$$

Using the above relations, we are ready to see

$$\begin{aligned} & p(r, U, \mathbf{X}, \mathbf{Y}, V, s) p(U) p(\mathbf{X}) p(\mathbf{Y}) p(V) \\ &= p(r, U) p(U, \mathbf{X}) p(\mathbf{X}, \mathbf{Y}) p(\mathbf{Y}, V) p(V, s), \end{aligned}$$

which matches an alternative definition of Markov chain given in [12, Section 2.1]. Other Markov chains shown in Fig. 1 can be verified accordingly.

IV. SYMMETRIC PROPERTIES AND SOME APPLICATIONS

We first state an intrinsic symmetric property of any LOCs, which induces other symmetric properties of LOCs used in this paper. A matrix is said to have full column (row) rank if its rank is equal to its number of columns (rows).

Theorem 1: For $\text{LOC}(H, T)$, if $\mathbf{X} = \mathbf{B}\mathbf{D}$ and $\mathbf{Y} = \mathbf{B}\mathbf{E}$, where \mathbf{B} has full column rank, then

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{\mathbf{X}H = \mathbf{Y}\} = \Pr\{\mathbf{D}H = \mathbf{E}\}.$$

Proof: The theorem follows from $P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{\mathbf{B}\mathbf{D}H = \mathbf{B}\mathbf{E}\} = \Pr\{\mathbf{D}H = \mathbf{E}\}$, where the last equality follows because \mathbf{B} has full column rank. ■

A. Computation of Transition Matrix

The matrix of transition probabilities is also called the transition matrix. Computing the transition matrix using (2) has a complexity $\mathcal{O}(q^{TM+MN})$ since we have q^{TM} choices of \mathbf{X} and for each \mathbf{X} , there are at most q^{MN} choices of \mathbf{Y} such that $P_{Y|X}(\mathbf{Y}|\mathbf{X}) \neq 0$. We can use Theorem 1 to reduce the complexity.

Let \mathbf{B} be a $t \times r$ matrix with rank r . For a $t \times m$ matrix \mathbf{A} with $\langle \mathbf{A} \rangle \subset \langle \mathbf{B} \rangle$, define \mathbf{A}/\mathbf{B} to be a matrix such that $\mathbf{A} = \mathbf{B}(\mathbf{A}/\mathbf{B})$. The notation “/” is well defined because i) there always exists \mathbf{C} such that $\mathbf{A} = \mathbf{B}\mathbf{C}$ since $\langle \mathbf{A} \rangle \subset \langle \mathbf{B} \rangle$ and ii) such \mathbf{C} is unique since \mathbf{B} is full column rank.

Corollary 2: Let \mathbf{X} and \mathbf{Y} be the input and output matrices of $\text{LOC}(H, T)$, respectively, with $\langle \mathbf{Y} \rangle \subset \langle \mathbf{X} \rangle$. Fix a full column rank matrix \mathbf{B} with $\langle \mathbf{X} \rangle = \langle \mathbf{B} \rangle$. Then,

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{(\mathbf{X}/\mathbf{B})H = \mathbf{Y}/\mathbf{B}\}. \quad (5)$$

Proof: Since $\mathbf{X} = \mathbf{B}(\mathbf{X}/\mathbf{B})$ and $\mathbf{Y} = \mathbf{B}(\mathbf{Y}/\mathbf{B})$, the result follows from Theorem 1. ■

In Corollary 2, the dimension of \mathbf{X}/\mathbf{B} is $\text{rk}(\mathbf{X}) \times M$ and the dimension of \mathbf{Y}/\mathbf{B} is $\text{rk}(\mathbf{X}) \times N$. Since $\text{rk}(\mathbf{X}) \leq M$ and \mathbf{X}/\mathbf{B} has full row rank, the computation of the transition matrix of $\text{LOC}(H, T)$ can be reduced to compute

$$\Pr\{\mathbf{D}H = \mathbf{E}\}, \quad \text{for all } \mathbf{D} \in \text{Fr}(\mathbb{F}^{k \times M}), \quad k = 0, 1, \dots, \min\{M, T\}, \quad (6)$$

where $\text{Fr}(\mathbb{F}^{k \times M})$ denotes the set of full rank matrices in $\mathbb{F}^{k \times M}$. For a fixed k , the number of \mathbf{D} needed to be considered is $|\text{Fr}(\mathbb{F}^{k \times M})|$, which is given by the number χ_k^M defined in (33) (see Lemma 8 in Appendix A).

We can further simplify the computation. For any $\mathbf{D} \in \text{Fr}(\mathbb{F}^{M \times M})$, by Corollary 2,

$$\Pr\{\mathbf{D}H = \mathbf{E}\} = \Pr\{H = \mathbf{E}\mathbf{D}^{-1}\}.$$

In other words, we only need to consider one matrix in $\text{Fr}(\mathbb{F}^{M \times M})$. The following symmetric property summarizes this observation.

Theorem 3: Consider a $\text{LOC}(H, T)$ where H has dimension $M \times N$. For $k \leq M$, $\mathbf{D}_1, \mathbf{D}_2 \in \text{Fr}(\mathbb{F}^{k \times M})$, if $\langle \mathbf{D}_1^\top \rangle = \langle \mathbf{D}_2^\top \rangle$, i.e., the row spaces of \mathbf{D}_1 and \mathbf{D}_2 are the same, the vector $(\Pr\{\mathbf{D}_1 H = \mathbf{E}\} : \mathbf{E} \in \mathbb{F}^{k \times N})$ is a permutation of the vector $(\Pr\{\mathbf{D}_2 H = \mathbf{E}\} : \mathbf{E} \in \mathbb{F}^{k \times N})$.

Proof: We only need to show that there exists a bijection $f : \mathbb{F}^{k \times N} \rightarrow \mathbb{F}^{k \times N}$ such that $\Pr\{\mathbf{D}_1 H = \mathbf{E}\} = \Pr\{\mathbf{D}_2 H = f(\mathbf{E})\}$. Since $\mathbf{D}_1, \mathbf{D}_2 \in \text{Fr}(\mathbb{F}^{k \times M})$ and $\langle \mathbf{D}_1^\top \rangle = \langle \mathbf{D}_2^\top \rangle$, there exists a unique full rank matrix \mathbf{T} such that $\mathbf{D}_2 = \mathbf{T}\mathbf{D}_1$. Define $f : \mathbb{F}^{k \times N} \rightarrow \mathbb{F}^{k \times N}$ as $f(\mathbf{E}) = \mathbf{T}\mathbf{E}$. Since \mathbf{T} is full rank, f is a bijection. The proof is completed by $\Pr\{\mathbf{D}_2 H = f(\mathbf{E})\} = \Pr\{\mathbf{D}_2 H = \mathbf{T}\mathbf{E}\} = \Pr\{\mathbf{T}^{-1}\mathbf{D}_2 H = \mathbf{E}\} = \Pr\{\mathbf{D}_1 H = \mathbf{E}\}$. ■

Let $\text{Pj}(m, \mathbb{F}^t)$ be the subset of $\text{Pj}(\mathbb{F}^t)$ that contains all the subspaces with dimension less than or equal to m . By Theorem 3, for each subspace in $\text{Pj}(\min\{T, M\}, \mathbb{F}^M)$, we only need to choose one \mathbf{D} to compute $(\Pr\{\mathbf{D}\mathbf{H} = \mathbf{E}\}, \mathbf{E} \in \mathbb{F}^{k \times N})$. The *Grassmannian* $\text{Gr}(r, \mathbb{F}^t)$ is the set of all r -dimensional subspaces of \mathbb{F}^t . Thus $\text{Pj}(\min\{T, M\}, \mathbb{F}^M) = \bigcup_{k \leq \min\{T, M\}} \text{Gr}(k, \mathbb{F}^M)$. By Lemma 10 in Appendix A, $|\text{Gr}(k, \mathbb{F}^M)| = \binom{M}{k}_q$, where $\binom{M}{k}_q$ is called the Gaussian binomials. So the overall complexity of computing the transition matrix is

$$\sum_{k=0}^{\min\{T, M\}} \binom{M}{k}_q q^{kN} < \begin{cases} cMq^{\left(\frac{M+N}{2}\right)^2}, & N \leq M, M \leq T \\ cMq^{MN}, & N > M, M \leq T \\ cTq^{\left(\frac{M+N}{2}\right)^2}, & T > \frac{M+N}{2}, M > T \\ cTq^{(M+N-T)T}, & T \leq \frac{M+N}{2}, M > T \end{cases},$$

where the $c \approx 0.3427$ is a constant (see Lemma 14 in Appendix B).

B. α -type Input Distributions

In general, accurately finding an optimal input distribution needs to determine q^{TM} probability masses. Here we show that the problem can be reduced to find an optimal distribution over $\text{Pj}(\min\{T, M\}, \mathbb{F}^M)$.

Definition 1: A PMF p over $\mathbb{F}^{T \times M}$ is α -type if $p(\mathbf{X}) = p(\mathbf{X}')$ for all $\mathbf{X}, \mathbf{X}' \in \mathbb{F}^{T \times M}$ with $\langle \mathbf{X}^\top \rangle = \langle \mathbf{X}'^\top \rangle$, i.e., the same row spaces.

Lemma 1: A function $p : \mathbb{F}^{T \times M} \rightarrow \mathbb{R}$ is an α -type PMF if and only if it can be written as

$$p(\mathbf{X}) = Q(\langle \mathbf{X}^\top \rangle) / \chi_{\text{rk}(\mathbf{X})}^T \quad (7)$$

for certain PMF Q over $\text{Pj}(\min\{M, T\}, \mathbb{F}^M)$.

Proof: Assume p is an α -type input. Define $Q : \text{Pj}(\min\{M, T\}, \mathbb{F}^M) \rightarrow \mathbb{R}$ as

$$Q(\tilde{U}) = \sum_{\mathbf{X}' \in \mathbb{F}^{T \times M} : \langle \mathbf{X}'^\top \rangle = \tilde{U}} p(\mathbf{X}').$$

For $\mathbf{X} \in \mathbb{F}^{T \times M}$,

$$\begin{aligned} Q(\langle \mathbf{X}^\top \rangle) &= \sum_{\mathbf{X}' \in \mathbb{F}^{T \times M} : \langle \mathbf{X}'^\top \rangle = \langle \mathbf{X}^\top \rangle} p(\mathbf{X}') \\ &= p(\mathbf{X}) \sum_{\mathbf{X}' \in \mathbb{F}^{T \times M} : \langle \mathbf{X}'^\top \rangle = \langle \mathbf{X}^\top \rangle} 1 \\ &= p(\mathbf{X}) \chi_{\text{rk}(\mathbf{X})}^T, \end{aligned}$$

where the last equality follows from Lemma 17 in Appendix B. This proves the necessary condition.

Now we prove the sufficient condition. Let Q be a PMF over $\text{Pj}(\min\{M, T\}, \mathbb{F}^M)$. Define a function $p : \mathbb{F}^{T \times M} \rightarrow \mathbb{R}$ as

$$p(\mathbf{X}) = Q(\langle \mathbf{X}^\top \rangle) / \chi_{\text{rk}(\mathbf{X})}^T.$$

We can check that for $\mathbf{X}, \mathbf{X}' \in \mathbb{F}^{T \times M}$ with $\langle \mathbf{X}^\top \rangle = \langle \mathbf{X}'^\top \rangle$,

$$\begin{aligned} p(\mathbf{X}) &= Q(\langle \mathbf{X}^\top \rangle) / \chi_{\text{rk}(\mathbf{X})}^T \\ &= Q(\langle \mathbf{X}'^\top \rangle) / \chi_{\text{rk}(\mathbf{X})}^T \\ &= p(\mathbf{X}'), \end{aligned}$$

and

$$\begin{aligned} \sum_{\mathbf{X}} p(\mathbf{X}) &= \sum_{\tilde{U} \in \text{Pj}(\mathbb{F}^M)} \sum_{\mathbf{X}: \langle \mathbf{X}^\top \rangle = \tilde{U}} Q(\tilde{U}) / \chi_{\dim(\tilde{U})}^T \\ &= \sum_{\tilde{U} \in \text{Pj}(\mathbb{F}^M)} Q(\tilde{U}) / \chi_{\dim(\tilde{U})}^T \sum_{\mathbf{X}: \langle \mathbf{X}^\top \rangle = \tilde{U}} 1 \\ &= \sum_{\tilde{U} \in \text{Pj}(\mathbb{F}^M)} Q(\tilde{U}) \\ &= 1. \end{aligned}$$

Thus p is an α -type PMF. ■

Theorem 4: There exists an α -type input that maximizes $I(X; Y)$ for any LOC, i.e.,

$$C(H, T) = \max_{p_X: \alpha\text{-type}} I(X; Y).$$

Proof: This theorem is proved using Theorem 1 and the concavity of mutual information as a function of input distribution. See Section IV-C for details. ■

Theorem 4 narrows down the range to find an optimal input. To determine a PMF over $\text{Pj}(\min\{M, T\}, \mathbb{F}^M)$, we have $|\text{Pj}(\min\{M, T\}, \mathbb{F}^M)|$ parameters to determine. We know $|\text{Pj}(\min\{M, T\}, \mathbb{F}^M)| = \sum_{k=0}^{\min\{M, T\}} \binom{M}{k}_q < c \min\{M, T\} q^{M^2/4}$, where $c \approx 3.4627$ (see Lemma 14 in Appendix B).

C. More Symmetric Properties

The following lemma is used to prove Theorem 4.

Lemma 2: Let p_X be an input distribution of $\text{LOC}(H, T)$ where H has dimension $M \times N$. Define $p'_X: \mathbb{F}^{T \times M} \rightarrow \mathbb{R}$ as $p'_X(\mathbf{X}) = p_X(\Phi \mathbf{X})$, where $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$. We have, i) p'_X is a PMF, ii) $I(X; Y)|_{p_X} = I(X; Y)|_{p'_X}$ and iii) $I(\langle X \rangle; \langle Y \rangle)|_{p_X} = I(\langle X \rangle; \langle Y \rangle)|_{p'_X}$.

Proof: First p'_X is a PMF because $0 \leq p'_X(\mathbf{X}) = p(\Phi \mathbf{X}) \leq 1$ and

$$\begin{aligned} \sum_{\mathbf{X} \in \mathbb{F}^{T \times M}} p'_X(\mathbf{X}) &= \sum_{\mathbf{X} \in \mathbb{F}^{T \times M}} p(\Phi \mathbf{X}) \\ &= \sum_{\mathbf{X} \in \Phi \mathbb{F}^{T \times M}} p(\mathbf{X}) \\ &= \sum_{\mathbf{X} \in \mathbb{F}^{T \times M}} p(\mathbf{X}) \\ &= 1. \end{aligned}$$

Let p_Y and p'_Y be the PMF of Y when the inputs are p_X and p'_X , respectively. We have

$$\begin{aligned}
 p'_Y(\mathbf{Y}) &= \sum_{\mathbf{X} \in \mathbb{F}^{T \times M}} p'_X(\mathbf{X}) P_{Y|X}(\mathbf{Y}|\mathbf{X}) \\
 &\stackrel{(a)}{=} \sum_{\mathbf{X} \in \mathbb{F}^{T \times M}} p(\Phi\mathbf{X}) P_{Y|X}(\Phi\mathbf{Y}|\Phi\mathbf{X}) \\
 &\stackrel{(b)}{=} \sum_{\mathbf{X}' \in \mathbb{F}^{T \times M}} p(\mathbf{X}') P_{Y|X}(\Phi\mathbf{Y}|\mathbf{X}') \\
 &= p_Y(\Phi\mathbf{Y}),
 \end{aligned}$$

where (a) follows from Theorem 1 and $p'_X(\mathbf{X}) = p_X(\Phi\mathbf{X})$, and (b) follows by letting $\mathbf{X}' = \Phi\mathbf{X}$. Therefore,

$$\begin{aligned}
 I(X; Y)|_{p'_X} &= \sum_{\mathbf{X}} p'_X(\mathbf{X}) \sum_{\mathbf{Y}} P(\mathbf{Y}|\mathbf{X}) \log_2 \frac{P(\mathbf{Y}|\mathbf{X})}{p'_Y(\mathbf{Y})} \\
 &\stackrel{(c)}{=} \sum_{\mathbf{X}} p(\Phi\mathbf{X}) \sum_{\mathbf{Y}} P(\Phi\mathbf{Y}|\Phi\mathbf{X}) \log_2 \frac{P(\Phi\mathbf{Y}|\Phi\mathbf{X})}{p(\Phi\mathbf{Y})} \\
 &= \sum_{\mathbf{X}'} p(\mathbf{X}') \sum_{\mathbf{Y}'} P(\mathbf{Y}'|\mathbf{X}') \log_2 \frac{P(\mathbf{Y}'|\mathbf{X}')}{p(\mathbf{Y}')} \\
 &= I(X; Y)|_{p_X},
 \end{aligned}$$

where (c) follows from Theorem 1.

The last equality in the lemma can be proved similarly. First,

$$\begin{aligned}
 p'_{\langle X \rangle}(U) &= \sum_{\mathbf{X}: \langle \mathbf{X} \rangle = U} p'_X(\mathbf{X}) \\
 &= \sum_{\mathbf{X}: \langle \mathbf{X} \rangle = U} p_X(\Phi\mathbf{X}) \\
 &\stackrel{(d)}{=} \sum_{\mathbf{X}: \langle \mathbf{X} \rangle = \Phi U} p_X(\mathbf{X}') \\
 &= p_{\langle X \rangle}(\Phi U),
 \end{aligned}$$

where (d) follows from Theorem 1. Let $P'_{\langle Y \rangle|\langle X \rangle}(V|U)$ be the transition probability when the input is p'_X . For $U \leq \mathbb{F}^T$ with $p_{\langle X \rangle}(U) > 0$,

$$\begin{aligned}
 &P'_{\langle Y \rangle|\langle X \rangle}(V|U) \\
 &= \frac{\sum_{\mathbf{X}, \mathbf{Y}: \langle \mathbf{X} \rangle = U, \langle \mathbf{Y} \rangle = V} P_{Y|X}(\mathbf{Y}|\mathbf{X}) p'_X(\mathbf{X})}{p'_{\langle X \rangle}(U)} \\
 &= \frac{\sum_{\mathbf{X}, \mathbf{Y}: \langle \mathbf{X} \rangle = U, \langle \mathbf{Y} \rangle = V} P_{Y|X}(\Phi\mathbf{Y}|\Phi\mathbf{X}) p_X(\Phi\mathbf{X})}{p_{\langle X \rangle}(\Phi U)} \\
 &= P_{\langle Y \rangle|\langle X \rangle}(\Phi V|\Phi U).
 \end{aligned}$$

Hence,

$$\begin{aligned}
p'_{\langle Y \rangle}(V) &= \sum_U P'_{\langle Y \rangle | \langle X \rangle}(V|U) p'_{\langle X \rangle}(U) \\
&= \sum_U P_{\langle Y \rangle | \langle X \rangle}(\Phi V | \Phi U) p_{\langle X \rangle}(\Phi U) \\
&= p_{\langle Y \rangle}(\Phi V).
\end{aligned}$$

Therefore,

$$\begin{aligned}
I(\langle X \rangle; \langle Y \rangle)_{|p'_X} &= \sum_U p'_{\langle X \rangle}(U) \sum_V P'(V|U) \log_2 \frac{P'(V|U)}{P'_{\langle Y \rangle}(V)} \\
&= \sum_U p_{\langle X \rangle}(\Phi U) \sum_V P(\Phi V | \Phi U) \log_2 \frac{P(\Phi V | \Phi U)}{p_{\langle Y \rangle}(\Phi V)} \\
&= I(\langle X \rangle; \langle Y \rangle)_{|p_X}.
\end{aligned}$$

■

Proof of Theorem 4: Consider $\text{LOC}(H, T)$. Let p be an optimal input distribution for the channel. For $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$, define p^Φ as $p^\Phi(\mathbf{X}) = p(\Phi \mathbf{X})$. By Lemma 2, $p^\Phi(\mathbf{X})$ also achieves the capacity of the LOC. Define p^* as

$$p^*(\mathbf{X}) = \frac{1}{|\text{Fr}(\mathbb{F}^{T \times T})|} \sum_{\Phi \in \text{Fr}(\mathbb{F}^{T \times T})} p^\Phi(\mathbf{X}).$$

By the concavity of the mutual information, we know that p^* is also an optimal input for the channel.

Now we show that p^* is α -type. Consider $\mathbf{X}, \mathbf{X}' \in \mathbb{F}^{T \times M}$ with $\langle \mathbf{X}^\top \rangle = \langle \mathbf{X}'^\top \rangle$. There always exists $\Phi_0 \in \text{Fr}(\mathbb{F}^{T \times T})$ such that $\mathbf{X}' = \Phi_0 \mathbf{X}$ (see Lemma 16 in Appendix B). We have

$$\begin{aligned}
p^*(\Phi_0 \mathbf{X}) &= \frac{1}{|\text{Fr}(\mathbb{F}^{T \times T})|} \sum_{\Phi \in \text{Fr}(\mathbb{F}^{T \times T})} p^\Phi(\Phi_0 \mathbf{X}) \\
&= \frac{1}{|\text{Fr}(\mathbb{F}^{T \times T})|} \sum_{\Phi \in \text{Fr}(\mathbb{F}^{T \times T})} p^{\Phi \Phi_0}(\mathbf{X}) \\
&= p^*(\mathbf{X}),
\end{aligned}$$

where in the last equality we use $\text{Fr}(\mathbb{F}^{T \times T}) = \Phi_0 \text{Fr}(\mathbb{F}^{T \times T})$. ■

V. SUBSPACE DEGRADATIONS OF LOCs

Consider the Markov chain $\langle X \rangle \rightarrow X \rightarrow Y \rightarrow \langle Y \rangle$ related to $\text{LOC}(H, T)$. The transition probability from X to Y is given by (2). The transition probability from Y to $\langle Y \rangle$ is deterministic:

$$P_{\langle Y \rangle | Y}(V | \mathbf{Y}) = \begin{cases} 1, & \langle \mathbf{Y} \rangle = V \\ 0, & o.w. \end{cases}$$

The transition probability $P_{X | \langle X \rangle}$ is not determined by the LOC.

Definition 2: Consider $\text{LOC}(H, T)$ with transition probability $P_{Y|X}$. Given a transition probability $P_{X|\langle X \rangle}$, we have a new channel law given by $P_{\langle Y \rangle|\langle X \rangle}(V|U)$. This channel takes subspaces as input and output and is called the subspace degradation of $\text{LOC}(H, T)$ with respect to (w.r.t) $P_{X|\langle X \rangle}$. This degradation is well defined since the transition probability $P_{\langle Y \rangle|\langle X \rangle}$ is determined by the above Markov chain as

$$\begin{aligned} P_{\langle Y \rangle|\langle X \rangle}(V|U) &= \sum_{\mathbf{X}} P_{\langle Y \rangle|X}(V|\mathbf{X}) P_{X|\langle X \rangle}(\mathbf{X}|U) \\ &= \sum_{\mathbf{X}: \langle \mathbf{X} \rangle = U} \sum_{\mathbf{Y}: \langle \mathbf{Y} \rangle = V} P_{Y|X}(\mathbf{Y}|\mathbf{X}) P_{X|\langle X \rangle}(\mathbf{X}|U). \end{aligned} \quad (8)$$

For a subspace degradation of $\text{LOC}(H, T)$ w.r.t $P_{X|\langle X \rangle}$, the mutual information between $\langle X \rangle$ and $\langle Y \rangle$ can be written as a function of $p_{\langle X \rangle}$ and $P_{\langle Y \rangle|\langle X \rangle}$, in which $P_{\langle Y \rangle|\langle X \rangle}$, given in (8), is a function of $P_{X|\langle X \rangle}(\mathbf{X}|U)$. The capacity of a subspace degradation of a LOC is $\max_{p_{\langle X \rangle}} I(\langle Y \rangle; \langle X \rangle)$. Therefore, the maximum achievable rate of subspace degradations of $\text{LOC}(H, T)$ is

$$C_{\text{ss}}(H, T) = \max_{p_{X|\langle X \rangle}} \max_{p_{\langle X \rangle}} I(\langle X \rangle; \langle Y \rangle).$$

The rate $C_{\text{ss}}(H, T)$ is achievable since $\max_{p_{\langle X \rangle}} I(\langle X \rangle; \langle Y \rangle)$ is achievable for any given $p_{X|\langle X \rangle}$.

Lemma 3: For $\text{LOC}(H, T)$, $I(\langle X \rangle; \langle Y \rangle)$ is determined by p_X , i.e., we can treat $I(\langle X \rangle; \langle Y \rangle)$ as a function of p_X for a given LOC and write

$$C_{\text{ss}}(H, T) = \max_{p_X} I(\langle X \rangle; \langle Y \rangle).$$

Proof: We show that $p_{\langle X \rangle}(U)$ and $P_{X|\langle X \rangle}(\mathbf{X}|U)$ appeared in $I(\langle X \rangle; \langle Y \rangle)$ are determined by p_X . First, we obtain $p_{\langle X \rangle}$ from p_X as shown in (3). Second, since

$$\begin{aligned} P_{X|\langle X \rangle}(\mathbf{X}|U) p_{\langle X \rangle}(U) &= \Pr\{X = \mathbf{X}, \langle X \rangle = U\} \\ &= \begin{cases} p_X(\mathbf{X}) & \langle \mathbf{X} \rangle = U \\ 0 & o.w. \end{cases}, \end{aligned}$$

we have

$$P_{X|\langle X \rangle}(\mathbf{X}|U) = \begin{cases} \frac{p_X(\mathbf{X})}{p_{\langle X \rangle}(U)} & p_{\langle X \rangle}(U) \neq 0, \langle \mathbf{X} \rangle = U \\ 0 & \langle \mathbf{X} \rangle \neq U. \end{cases} \quad (9)$$

That means, for U with $p_{\langle X \rangle}(U) > 0$, $P_{X|\langle X \rangle}(\mathbf{X}|U)$ is determined by p_X . Moreover, if $p_{\langle X \rangle}(U) = 0$, $P_{X|\langle X \rangle}(\mathbf{X}|U)$ does not appear in $I(\langle X \rangle; \langle Y \rangle)$. Thus, $I(\langle X \rangle; \langle Y \rangle)$ can be regarded as a function with only one variable p_X . This also implies that

$$C_{\text{ss}}(H, T) \geq \max_{p_X} I(\langle X \rangle; \langle Y \rangle).$$

One the other hand, given $P_{X|\langle X \rangle}$ and $p_{\langle X \rangle}$, we have a PMF of X given by

$$p_X(\mathbf{X}) = p_{\langle X \rangle}(\langle \mathbf{X} \rangle) P_{X|\langle X \rangle}(\mathbf{X}|\langle \mathbf{X} \rangle),$$

which establishes that

$$C_{\text{ss}}(H, T) \leq \max_{p_X} I(\langle X \rangle; \langle Y \rangle).$$

The proof is completed. ■

A. Uniform LOCs

In general $I(X; Y) \geq I(\langle X \rangle; \langle Y \rangle)$. We may want to know when the equality holds, under which using subspaces suffices.

Definition 3: $\text{LOC}(H, T)$ is *uniform (for given input and output subspaces)* if there exists a function $\mu : \text{Pj}(\mathbb{F}^T) \times \text{Pj}(\mathbb{F}^T) \rightarrow [0, 1]$ such that

$$\Pr\{\mathbf{Y} = \mathbf{X}H\} = \begin{cases} \mu(\langle \mathbf{X} \rangle, \langle \mathbf{Y} \rangle) & \langle \mathbf{Y} \rangle \subseteq \langle \mathbf{X} \rangle \\ 0 & \text{o.w.} \end{cases}$$

Theorem 5: A uniform $\text{LOC}(H, T)$ where H has dimension $M \times N$ has a unique subspace degradation given by

$$P_{\langle Y \rangle | \langle X \rangle}(V|U) = \mu(U, V) \chi_{\dim(V)}^N.$$

Moreover, for a uniform LOC, $I(X; Y) = I(\langle X \rangle; \langle Y \rangle)$.

Proof: See Section V-B. ■

The number of $M \times N$ matrices with rank r is given by $\chi_r^{M, N}$ defined in (36) (see Lemma 11 in Appendix A). There is another property such that $I(X; Y) = I(\langle X \rangle; \langle Y \rangle)$ holds.

Definition 4: A random matrix H over $\mathbb{F}^{M \times N}$ is *uniform (for a given rank)* if

$$p_H(\mathbf{H}) = \frac{p_{\text{rk}(H)}(\text{rk}(\mathbf{H}))}{\chi_{\text{rk}(\mathbf{H})}^{M, N}}.$$

Theorem 6: Let H be a random matrix with dimension $M \times N$. i) If $\text{LOC}(H, T)$ is uniform and $T \geq M$, then H is uniform. ii) If H is uniform, then $\text{LOC}(H, T)$ is uniform.

Proof: See Section VIII-C. ■

Now we give a uniform LOC that has a non-uniform transfer matrix. Let H be a 2×2 random matrix over the binary field with

$$p_H \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) = p_H \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) = p_H \left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = p_H \left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right) = 0.25.$$

H is not uniform in the sense of Definition 4, but we can verify that $\text{LOC}(H, 1)$ is uniform in the sense of Definition 3.

B. Proof of Theorem 5

Consider a subspace degradation of $\text{LOC}(H, T)$ with $P_{X|\langle X \rangle}(\mathbf{X}|U)$. First,

$$\begin{aligned} P_{Y|\langle X \rangle}(\mathbf{Y}|U) &= \sum_{\mathbf{X}: \langle \mathbf{X} \rangle = U} P(\mathbf{Y}|\mathbf{X})P(\mathbf{X}|U) \\ &= \sum_{\mathbf{X}: \langle \mathbf{X} \rangle = U} \mu(U, \langle \mathbf{Y} \rangle)P(\mathbf{X}|U) \\ &= \mu(U, \langle \mathbf{Y} \rangle). \end{aligned}$$

Then,

$$\begin{aligned} P_{\langle Y \rangle | \langle X \rangle}(V|U) &= \sum_{\mathbf{Y}: \langle \mathbf{Y} \rangle = V} P_{Y|X}(\mathbf{Y}|U) \\ &= \mu(U, V) \chi_{\dim(V)}^N. \end{aligned}$$

So this means that all $P_{X|\langle X \rangle}(\mathbf{X}|U)$ give the same subspace degradation.

Now we prove the rest part of the theorem. Let $\mathcal{U} = \text{Pj}(\mathbb{F}^T)$. We have

$$\begin{aligned} I(X; Y) &= \sum_{\mathbf{X}, \mathbf{Y}} p_{X,Y}(\mathbf{X}, \mathbf{Y}) \log_2 \frac{p_{X,Y}(\mathbf{X}, \mathbf{Y})}{p_X(\mathbf{X})p_Y(\mathbf{Y})} \\ &= \sum_{V, U \in \mathcal{U}} \sum_{\substack{\mathbf{X}, \mathbf{Y}: \\ \langle \mathbf{X} \rangle = U, \langle \mathbf{Y} \rangle = V}} p(\mathbf{X}, \mathbf{Y}) \log_2 \frac{p(\mathbf{X}, \mathbf{Y})}{p_X(\mathbf{X})p_Y(\mathbf{Y})} \\ &\leq \sum_{V, U \in \mathcal{U}} p_{\langle X \rangle \langle Y \rangle}(U, V) \log_2 \frac{p_{\langle X \rangle \langle Y \rangle}(U, V)}{p_{\langle X \rangle}(U)p_{\langle Y \rangle}(V)} \\ &= I(\langle X \rangle; \langle Y \rangle), \end{aligned} \tag{10}$$

where (10) follows from the log-sum inequality. To prove this theorem, we only need to show the equality in (10) holds for uniform LOCs. We need to check that $P_{Y|X}(\mathbf{Y}|\mathbf{X})/p_Y(\mathbf{Y})$ is a constant for all \mathbf{X} and \mathbf{Y} with $\langle \mathbf{Y} \rangle = V \leq \langle \mathbf{X} \rangle = U \leq \mathbb{F}^T$. Fix an input distribution p_X . Since the LOC is uniform,

$$\begin{aligned} p_Y(\mathbf{Y}) &= \sum_{\mathbf{X}: V \leq \langle \mathbf{X} \rangle} P_{Y|X}(\mathbf{Y}|\mathbf{X})p_X(\mathbf{X}) \\ &= \sum_{U' \leq \mathbb{F}^T: V \leq U'} \mu(U, V) \sum_{\mathbf{X}: \langle \mathbf{X} \rangle = U'} p_X(\mathbf{X}) \\ &= \sum_{U' \leq \mathbb{F}^T: V \leq U'} \mu(U', V) p_{\langle X \rangle}(U'). \end{aligned}$$

Thus,

$$\frac{P_{Y|X}(\mathbf{Y}|\mathbf{X})}{p_Y(\mathbf{Y})} = \frac{\mu(U, V)}{\sum_{U' \leq \mathbb{F}^T: V \leq U'} \mu(U', V) p_{\langle X \rangle}(U')}.$$

This verifies the equality in () holding.

VI. MUTUAL INFORMATION DECOMPOSITION

For subspace degradations, α -type input distributions are also useful.

Theorem 7: There exists an α -type input that maximizes $I(\langle X \rangle; \langle Y \rangle)$ for any LOC, i.e.,

$$C_{\text{ss}}(H, T) = \max_{p_X: \alpha\text{-type}} I(\langle X \rangle; \langle Y \rangle).$$

Proof: This theorem can be proved similar to Theorem 4 by applying Lemma 2. ■

For a random matrix X , recall that $\text{rk}(X)$ is the random variable representing the rank of X (see (4) for the PMF). Similar to Lemma 3, $I(\text{rk}(X); \text{rk}(Y))$ is determined by p_X and $P_{Y|X}$. Define

$$\begin{aligned} J(\text{rk}(X); \text{rk}(Y)) &= \sum_{s,r} p_{\text{rk}(X) \text{rk}(Y)}(r, s) \log_2 \frac{\chi_s^T}{\chi_s^r} \\ &= \mathbb{E} \left[\log_2 \frac{\chi_{\text{rk}(Y)}^T}{\chi_{\text{rk}(X)}^r} \right], \end{aligned} \quad (11)$$

where $p_{\text{rk}(X) \text{rk}(Y)}(r, s)$ can be derived using p_X and $P_{Y|X}$.

Theorem 8: For a LOC with α -type inputs,

$$I(\langle X \rangle; \langle Y \rangle) = I(\text{rk}(X); \text{rk}(Y)) + J(\text{rk}(X); \text{rk}(Y)). \quad (12)$$

Proof: The proof is done by rewriting the formulation of mutual information using the symmetric property and the definition of α -type inputs. See Section VI-B for details. ■

In (12), $I(\text{rk}(X); \text{rk}(Y))$ is the mutual information of the ranks of transmitted and received matrices. In other words, it is the rate transmitted using the matrix ranks. The meaning of $J(\text{rk}(X); \text{rk}(Y))$ has an interpretation using set packing. The capacity contributed by r -dimensional transmissions and s -dimensional receptions is $\log_2 \frac{\chi_s^T}{\chi_s^r} = \log_2 \binom{T}{s}_q / \binom{r}{s}_q$, where $\binom{T}{s}_q$ is the total number of s -dimensional subspaces in \mathbb{F}^T , and $\binom{r}{s}_q$ is the total number of s -dimensional subspaces in an r -dimensional subspace. Treat an s -dimensional subspace in \mathbb{F}^T as a set element. An r -dimension transmission can be regarded as a collection of s dimensional subspaces that span it. Then, the *maximum set packing* problem is looking for the maximum number of pairwise disjoint collections of s -dimensional subspaces which have cardinality $\binom{M}{r}_q$ and span an M -dimensional subspace.

One simple coding scheme of LOCs is to use part of X to recover the instance of H in the receiver. Such a scheme is referred as channel training and can only achieve rate $(1 - M/T) \mathbb{E}[\text{rk}(H)]$ (see the analysis in [13]). Theorem 8 implies that using subspace coding can achieve a rate strictly higher than using channel training.

Corollary 9: For $\text{LOC}(H, T)$ where H has dimension $M \times N$ and $T \geq M$,

$$\begin{aligned} C_{\text{ss}}(H, T) &\geq \mathbb{E} \left[\log_2 \frac{\chi_{\text{rk}(H)}^T}{\chi_{\text{rk}(H)}^M} \right] \\ &= (T - M) \mathbb{E}[\text{rk}(H)] \log_2 q + \epsilon(T, q), \end{aligned} \quad (13)$$

where $0 < \epsilon(T, q) = \sum_s p_{\text{rk}(H)}(s) \log_2 \frac{\zeta_s^T}{\zeta_s^M} < 1.8$. This lower bound is achieved by the α -type input p_X with $p_{\text{rk}(X)}(M) = 1$.

Proof: See Section VI-B. ■

Remark: Note that this bound depends on the rank distribution of the transformation matrix. This lower bound is in fact tight for certain LOCs with sufficiently large T (see Theorem 13).

A. A Useful Form

Lemma 4: Let \mathbf{X} be an input matrix of $\text{LOC}(H, T)$. Then,

$$P_{\text{rk}(Y)|X}(s|\mathbf{X}) = P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\langle \mathbf{X}^\top \rangle) = \Pr\{\text{rk}(\mathbf{D}H) = s\},$$

where \mathbf{D} is any $\text{rk}(\mathbf{X}) \times M$ matrix with $\langle \mathbf{D}^\top \rangle = \langle \mathbf{X}^\top \rangle$.

Proof: See Section VI-B. ■

We have a refined version of Lemma 1.

Lemma 5: A function $p : \mathbb{F}^{T \times M} \rightarrow \mathbb{R}$ is an α -type PMF if and only if it can be written as

$$p(\mathbf{X}) = R(\text{rk}(\mathbf{X})) \frac{Q_{\text{rk}(\mathbf{X})}(\langle \mathbf{X}^\top \rangle)}{\chi_{\text{rk}(\mathbf{X})}^T} \quad (14)$$

where $Q_r(\cdot)$ is a PMF over $\text{Gr}(r, \mathbb{F}^M)$ and $R(\cdot)$ be a PMF over $\{0, 1, \dots, M\}$.

Proof: If p can be written as (14), by Lemma 1, p is an α -type PMF. On the other hand, if p is an α -type PMF, it can be written as (7). Let

$$R(r) = \sum_{\tilde{U} : \text{rk}(\tilde{U})=r} Q(\tilde{U}).$$

For r such that $R(r) > 0$, let

$$Q_r(\tilde{U}) = \begin{cases} Q(\tilde{U})/R(r) & \text{dim}(\tilde{U}) = r \\ 0 & \text{o.w.} \end{cases}$$

For r such that $R(r) = 0$, let $Q_r(\cdot)$ be any PMF over $\text{Gr}(r, \mathbb{F}^M)$. Since $Q_{\text{dim}(\tilde{U})}(\tilde{U})R(\text{dim}(\tilde{U})) = Q(\tilde{U})$, we see that p can be written as (14). ■

When using the formulation in (14), $I(\text{rk}(X); \text{rk}(Y))$ and $J(\text{rk}(X); \text{rk}(Y))$ can be written as functions of $Q_r(\tilde{U})$ and $R(r)$ as follows. Using the property of Markov chain,

$$\begin{aligned} P_{\text{rk}(Y)|\text{rk}(X)}(s|r) &= \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) P_{\langle X^\top \rangle|\text{rk}(X)}(\tilde{U}|r) \\ &= \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) Q_r(\tilde{U}), \end{aligned} \quad (15)$$

in which $P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U})$, given in Lemma 4, is a function of p_H and is not related to $Q_r(\tilde{U})$ and $R(r)$. Thus, we can write

$$I(\text{rk}(X); \text{rk}(Y)) = \sum_r R(r) \sum_s P(s|r) \log_2 \frac{P(s|r)}{\sum_{r'} P(s|r') R(r')}, \quad (16)$$

where $P(s|r) = P_{\text{rk}(Y)|\text{rk}(X)}(s|r)$ is given in (15). On the other hand,

$$J(\text{rk}(X); \text{rk}(Y)) = \sum_r R(r) \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} Q_r(\tilde{U}) g(\tilde{U}),$$

where

$$g(\tilde{U}) \triangleq \sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) \log_2 \frac{\chi_s^T}{\chi_s^{\text{dim}(\tilde{U})}}. \quad (17)$$

Note that $g(\tilde{U})$ depends on only the distribution of H but not the input.

B. Proofs

Proof of Theorem 8: Fix an α -type input p_X . For $V \leq U \leq \mathbb{F}^T$ with $\dim(U) = r$ and $\dim(V) = s$, we first show

$$p_{\langle X \rangle \langle Y \rangle}(U, V) = \frac{p_{\text{rk}(X) \text{rk}(Y)}(r, s)}{\binom{T}{r}_q \binom{T}{s}_q}. \quad (18)$$

We only need to show that $p_{\langle X \rangle \langle Y \rangle}(U, V) = p_{\langle X \rangle \langle Y \rangle}(U', V')$ for any $V \leq U \leq \mathbb{F}^T$ and $V' \leq U' \leq \mathbb{F}^T$ with $\dim(U) = \dim(U')$ and $\dim(V) = \dim(V')$, because if this is true,

$$\begin{aligned} p_{\text{rk}(X) \text{rk}(Y)}(r, s) &= \sum_{\dim(U^*)=r, \dim(V^*)=s, V^* \subset U^*} p_{\langle X \rangle \langle Y \rangle}(U^*, V^*) \\ &= p_{\langle X \rangle \langle Y \rangle}(U, V) \sum_{\dim(U^*)=r, \dim(V^*)=s, V^* \subset U^*} 1 \\ &= p_{\langle X \rangle \langle Y \rangle}(U, V) \binom{T}{r}_q \binom{T}{s}_q. \end{aligned}$$

Let

$$A(m, U) = \{\mathbf{X} \in \mathbb{F}^{T \times m} : \langle \mathbf{X} \rangle = U\}.$$

There exists $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$ such that $\Phi U = U'$ and $\Phi V = V'$ (see Lemma 15 in Appendix B). Then,

$$\begin{aligned} p_{\langle X \rangle \langle Y \rangle}(U, V) &= \sum_{\mathbf{X} \in A(M, U)} p_X(\mathbf{X}) \sum_{\mathbf{Y} \in A(N, V)} P_{Y|X}(\mathbf{Y}|\mathbf{X}) \\ &= \sum_{\mathbf{X} \in A(M, U)} p_X(\Phi \mathbf{X}) \sum_{\mathbf{Y} \in A(N, V)} P_{Y|X}(\Phi \mathbf{Y}|\Phi \mathbf{X}) \end{aligned} \quad (19)$$

$$\begin{aligned} &= \sum_{\mathbf{X} \in A(M, \Phi U)} p_X(\mathbf{X}) \sum_{\mathbf{Y} \in A(N, \Phi V)} P_{Y|X}(\mathbf{Y}|\mathbf{X}) \quad (20) \\ &= p_{\langle X \rangle \langle Y \rangle}(\Phi U, \Phi V) \\ &= p_{\langle X \rangle \langle Y \rangle}(U', V'), \end{aligned}$$

where (19) follows that p_X is α -type ($p_X(\mathbf{X}) = p_X(\Phi \mathbf{X})$) and $P_{Y|X}(\Phi \mathbf{Y}|\Phi \mathbf{X}) = P_{Y|X}(\mathbf{Y}|\mathbf{X})$ follows from Theorem 1, (20) follows from $A(m, \Phi U) = \Phi A(m, U)$ (see Lemma 17). This proves (18).

Applying the property of α -type inputs,

$$\begin{aligned} p_{\langle X \rangle}(U) &= \sum_{\mathbf{X} \in A(M, U)} p_X(\mathbf{X}) \\ &= \sum_{\mathbf{X} \in A(M, U)} p_X(\Phi \mathbf{X}) \\ &= \sum_{\mathbf{X} \in \Phi A(M, U)} p_X(\mathbf{X}) \\ &= \sum_{\mathbf{X} \in A(M, U')} p_X(\mathbf{X}) \quad (21) \\ &= p_{\langle X \rangle}(U') \end{aligned}$$

where (21) follows from Lemma 17. Therefore,

$$p_{\langle X \rangle}(U) = \frac{p_{\text{rk}(X)}(r)}{\binom{T}{r}_q}. \quad (22)$$

Moreover,

$$\begin{aligned} p_{\langle Y \rangle}(V) &= \sum_{U: V \subset U} p_{\langle X \rangle \langle Y \rangle}(U, V) \\ &= \sum_{r \geq s} \sum_{U: V \subset U, \dim(U)=r} p_{\langle X \rangle \langle Y \rangle}(U, V) \\ &= \sum_{r \geq s} \frac{p_{\text{rk}(X)} \text{rk}(Y)(r, s)}{\binom{T}{r}_q \binom{r}{s}_q} \sum_{U: V \subset U, \dim(U)=r} 1 \\ &= \sum_{r \geq s} \frac{p_{\text{rk}(X)} \text{rk}(Y)(r, s)}{\binom{T}{r}_q \binom{r}{s}_q} \binom{T-s}{r-s}_q \end{aligned} \quad (23)$$

$$= \sum_{r \geq s} \frac{p_{\text{rk}(X)} \text{rk}(Y)(r, s)}{\binom{T}{r}_q \binom{r}{s}_q} \binom{T}{r}_q \frac{\chi_s^r}{\chi_s^T} \quad (24)$$

$$= \frac{p_{\text{rk}(Y)}(s)}{\binom{T}{s}_q}, \quad (25)$$

where (23) and (24) follow from Lemma 12 in Appendix A. Substituting (18), (22) and (25) into $I(\langle X \rangle; \langle Y \rangle)$, we have

$$\begin{aligned} I(\langle X \rangle; \langle Y \rangle) &= \sum_{V \leq U} p_{\langle X \rangle \langle Y \rangle}(U, V) \log_2 \frac{p_{\langle X \rangle \langle Y \rangle}(U, V)}{p_{\langle X \rangle}(U) p_{\langle Y \rangle}(V)} \\ &= \sum_{s \leq r} \sum_{\substack{V \leq U, \dim(U)=r, \\ \dim(V)=s}} p_{\langle X \rangle \langle Y \rangle}(U, V) \log_2 \frac{p_{\langle X \rangle \langle Y \rangle}(U, V)}{p_{\langle X \rangle}(U) p_{\langle Y \rangle}(V)} \\ &= \sum_{s \leq r} \sum_{\substack{V \leq U, \dim(U)=r, \\ \dim(V)=s}} p_{\langle X \rangle \langle Y \rangle}(U, V) \log_2 \frac{p_{\text{rk}(X)} \text{rk}(Y)(r, s)}{p_{\text{rk}(X)}(r) p_{\text{rk}(Y)}(s)} \frac{\binom{T}{s}_q}{\binom{r}{s}_q} \\ &= \sum_{s \leq r} p_{\text{rk}(X)} \text{rk}(Y)(s, r) \log_2 \frac{p_{\text{rk}(X)} \text{rk}(Y)(r, s)}{p_{\text{rk}(X)}(r) p_{\text{rk}(Y)}(s)} \frac{\binom{T}{s}_q}{\binom{r}{s}_q} \\ &= I(\text{rk}(X); \text{rk}(Y)) + \sum_{s \leq r} p_{\text{rk}(X)} \text{rk}(Y)(r, s) \log_2 \frac{\chi_s^T}{\chi_s^r}. \end{aligned}$$

This completes the proof. ■

Proof of Corollary 9: Substituting the α -type input with $p_{\text{rk}(X)}(M) = 1$ in Theorem 8, we have $I(\text{rk}(X); \text{rk}(Y)) = 0$ and $J(\text{rk}(X); \text{rk}(Y)) = \sum_s P_{\text{rk}(Y)|\text{rk}(X)}(s|M) \log_2 \frac{\chi_s^T}{\chi_s^M}$. Given $\mathbf{X} \in \mathbb{F}^{T \times M}$ with dimension M ,

$$P_{\text{rk}(Y)|X}(s|\mathbf{X}) = \Pr\{\text{rk}(\mathbf{X}H) = s\} = \Pr\{\text{rk}(H) = s\}.$$

Thus, $P_{\text{rk}(Y)|\text{rk}(X)}(s|M) = \Pr\{\text{rk}(H) = s\}$. Using the definition in (34), we can write

$$\begin{aligned} \log_2 \frac{\chi_s^T}{\chi_s^M} &= \log_2 \frac{\zeta_s^T q^{Ts}}{\zeta_s^M q^{Ms}} \\ &= (T - M)s \log_2 q + \log_2 \frac{\zeta_s^T}{\zeta_s^M}. \end{aligned}$$

Since $\zeta_s^T < 1$,

$$\log_2 \frac{\zeta_s^T}{\zeta_s^M} < \log_2 \frac{1}{\zeta_s^M} < 1.8, \quad (26)$$

where the last inequality follows from Lemma 13 in Appendix B. So

$$\begin{aligned} J(\text{rk}(X); \text{rk}(Y)) &= \sum_s p_{\text{rk}(H)}(s) (T - M)s \log_2 q + \\ &\quad \sum_s p_{\text{rk}(H)}(s) \log_2 \frac{\zeta_s^T}{\zeta_s^M} \\ &= (T - M) \log_2 q \mathbb{E}[\text{rk}(H)] + \epsilon(T, q), \end{aligned}$$

where $\epsilon(T, q) = \sum_s p_{\text{rk}(H)}(s) \log_2 \frac{\zeta_s^T}{\zeta_s^M} < 1.8$. The proof is completed by $C_{\text{SS}}(H, T) \geq J(\text{rk}(X); \text{rk}(Y))$. \blacksquare

Proof of Lemma 4: Fix a $\text{rk}(\mathbf{X}) \times M$ matrix \mathbf{D} with $\langle \mathbf{X}^\top \rangle = \langle \mathbf{D}^\top \rangle$. Let $\mathbf{B}^\top = \mathbf{X}^\top / \mathbf{D}^\top$. We know \mathbf{B} has full column rank. Since $X \rightarrow Y \rightarrow \text{rk}(Y)$ forms a Markov chain,

$$\begin{aligned} P_{\text{rk}(Y)|X}(s|\mathbf{X}) &= \sum_{\mathbf{Y}} P_{\text{rk}(Y)|Y}(s|\mathbf{Y}) P_{Y|X}(\mathbf{Y}|\mathbf{X}) \\ &= \sum_{\mathbf{Y}: \text{rk}(\mathbf{Y})=s} P_{Y|X}(\mathbf{Y}|\mathbf{X}) \\ &= \sum_{\mathbf{Y}: \text{rk}(\mathbf{Y})=s} \Pr\{\mathbf{D}\mathbf{H} = \mathbf{Y}/\mathbf{B}\} \\ &= \sum_{\mathbf{E}: \text{rk}(\mathbf{E})=s} \Pr\{\mathbf{D}\mathbf{H} = \mathbf{E}\} \\ &= \Pr\{\text{rk}(\mathbf{D}\mathbf{H}) = s\}, \end{aligned} \quad (27)$$

where (27) follows from (5).

Let $\tilde{U} = \langle \mathbf{X}^\top \rangle$. By the Markov chain $\langle X^\top \rangle \rightarrow X \rightarrow \text{rk}(Y)$,

$$\begin{aligned} &P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) \\ &= \sum_{\mathbf{X}': \langle \mathbf{X}'^\top \rangle = \tilde{U}} P_{\text{rk}(Y)|X}(s|\mathbf{X}') P_{X|\langle X^\top \rangle}(\mathbf{X}'|\tilde{U}) \\ &= \Pr\{\text{rk}(\mathbf{D}\mathbf{H}) = s\} \sum_{\mathbf{X}': \langle \mathbf{X}'^\top \rangle = \tilde{U}} P_{X|\langle X^\top \rangle}(\mathbf{X}'|\tilde{U}) \\ &= \Pr\{\text{rk}(\mathbf{D}\mathbf{H}) = s\}. \end{aligned}$$

The proof is completed. \blacksquare

VII. CONSTANT-RANK INPUTS FOR SUBSPACE DEGRADATIONS

An input distribution with $p_{\text{rk}(X)}(r) = 1$ is called a *constant-rank or rank- r input distribution*. Note that for a subspace degradation, using rank- r input is corresponding to using r -dimensional subspace coding. Let

$$C_{\text{C-SS}}(H, T) = \max_{p_X : \text{constant-rank}} I(\langle X \rangle; \langle Y \rangle),$$

i.e., $C_{\text{C-SS}}(H, T)$ is the maximum achievable rate of constant-dimensional subspace coding. The rank of a constant-rank input that achieves $C_{\text{C-SS}}(H, T)$ is called the *optimal input rank*.

We will prove the following theorem in Section VII-C by a way similar to proving Theorem 4.

Theorem 10: There exists a constant-rank α -type input that achieves $C_{\text{C-SS}}(H, T)$ for any LOC.

Theorem 11: For $\text{LOC}(H, T)$ where H has dimension $M \times N$, let

$$U^* = \arg \max_{\tilde{U} \in \text{Pj}(\min\{M, T\}, \mathbb{F}^M)} g(\tilde{U}),$$

where $g(\tilde{U})$ is defined in (17). Then, $r^* = \dim(\tilde{U}^*)$ is an optimal input rank and $C_{\text{C-SS}}(H, T) = g(\tilde{U}^*)$. Furthermore,

$$C_{\text{SS}}(H, T) - C_{\text{C-SS}}(H, T) \leq \max_{p_X} I(\text{rk}(X); \text{rk}(Y)) \leq \log_2 \min\{M, N, T\}.$$

Proof: See Section VII-C. ■

A. Optimal Input Rank

For $\text{LOC}(H, T)$, define

$$\text{rk}^*(H) = \max\{r : \Pr\{\text{rk}(H) = r\} > 0\}.$$

Lemma 6: Consider $\text{LOC}(H, T)$ where H has dimension $M \times N$ and $T \geq M$. Fix an α -type input. For $\tilde{V} \leq \mathbb{F}^M$ with $\dim(\tilde{V}) = r < \text{rk}^*(H)$,

$$g(\mathbb{F}^M) - g(\tilde{V}) > \Theta(T, r, H) \log_2 q,$$

where

$$\begin{aligned} \Theta(T, r, H) &= (T - M)(\text{rk}^*(H) - r)p_{\text{rk}(H)}(\text{rk}^*(H)) \\ &\quad - r(M - r) + \log_q \zeta_r^T. \end{aligned}$$

Proof: See Section VII-C. ■

Theorem 12: For $\text{LOC}(H, T)$, there exists T_0 such that when $T \geq T_0$, $r^* \geq \text{rk}^*(H)$, where r^* is the optimal input rank given in Theorem 11.

Proof of Theorem 12: Suppose the dimension H is $M \times N$. Fix T_0 such that $\Theta(T_0, r, H) \geq 0$ for all $r < \text{rk}^*(H)$. This is possible because $\Theta(T, r, H)$ is a linearly increasing function of T for all $r < \text{rk}^*(H)$. Assume $T \geq T_0$ and $r^* < \text{rk}^*(H)$. For any $\tilde{V} \leq \mathbb{F}^M$ with $\dim(\tilde{V}) < \text{rk}^*(H) \leq M$, by Lemma 6, $g(\mathbb{F}^M) > g(\tilde{V})$. Thus, we have a contradiction to $r^* < \text{rk}^*(H)$. ■

Theorem 12 narrows down the range to search an optimal input rank for large T . The proof tells that there exists a T_0 as

$$T_0 = \min\{T : \Theta(T, r, H) \geq 0, r < \text{rk}^*(H)\}.$$

When $\text{rk}^*(H) = M$, we can get that

$$T_0 = M + \frac{M - 1 - \log_q \zeta_r^r}{p_{\text{rk}(H)}(M)}.$$

B. Optimality of Using Constant-Rank Inputs

For subspace degradations, we have shown that the loss of rate by using only constant-dimensional subspace coding is upper bounded. In fact, using constant-rank is optimal for subspace degradations under the following constraints.

Definition 5: A random matrix H with dimension $M \times N$ is *regular* if $p_{\text{rk}(H)}(s) > 0$ for $0 \leq s \leq M$. Furthermore, $\text{LOC}(H, T)$ is regular if H is regular.

Theorem 13: Consider regular $\text{LOC}(H, T)$ where H has dimension $M \times N$. There exists T_1 such that when $T \geq T_1$, C_{SS} is achieved by the α -type input with $R(M) = 1$. In this case $C_{\text{SS}}(H, T) = g(\mathbb{F}^M) = \sum_s p_{\text{rk}(H)}(s) \log_2 \frac{\chi_s^T}{\chi_s^M} = \mathbb{E} \left[\log_2 \frac{\chi_{\text{rk}(H)}^T}{\chi_{\text{rk}(H)}^M} \right]$.

Proof: See Section VII-C. ■

C. Proofs

Proof of Theorem 10: Consider a LOC with block length T . Let $p_X(\mathbf{X})$ be an optimal constant-rank input with $p_{\text{rk}(X)}(r^*) = 1$. For $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$, define p^Φ as $p_X^\Phi(\mathbf{X}) = p_X(\Phi \mathbf{X})$. It is clear that $p_{\text{rk}(X)}^\Phi(r^*) = 1$. By Lemma 2, $p_X^\Phi(\mathbf{X})$ is also an optimal constant-rank input. Define p_X^* as

$$p_X^*(\mathbf{X}) = \frac{1}{|\text{Fr}(\mathbb{F}^{T \times T})|} \sum_{\Phi \in \text{Fr}(\mathbb{F}^{T \times T})} p_X^\Phi(\mathbf{X}).$$

By the concavity of the mutual information, p_X^* is also an optimal constant-rank input. Similar to the procedure in the proof of Theorem 4, we can check that p_X^* is an α -type distribution. ■

Proof of Theorem 11: For an r -dimensional α -type input,

$$\begin{aligned} I(\langle X \rangle; \langle Y \rangle) &= \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} Q_r(\tilde{U}) g(\tilde{U}) \\ &\leq \max_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} g(\tilde{U}) \\ &\leq g(\tilde{U}^*). \end{aligned}$$

Thus $C_{\text{C-SS}} \leq g(\tilde{U}^*)$. On the other hand, for the r^* -dimensional α -type input with $p_{\langle X \rangle}(\tilde{U}^*) = 1$, $C_{\text{C-SS}} \geq I(\langle X \rangle; \langle Y \rangle) = g(\tilde{U}^*)$.

Furthermore, for an α -type input $I(\langle X \rangle; \langle Y \rangle) - C_{\text{C-SS}} = I(\text{rk}(X); \text{rk}(Y)) + J(\text{rk } X; \text{rk } Y) - g(\tilde{U}^*) \leq I(\text{rk}(X); \text{rk}(Y))$.

Thus, $C_{\text{SS}} - C_{\text{C-SS}} = \max_{p_X: \alpha\text{-type}} I(\langle X \rangle; \langle Y \rangle) - C_{\text{C-SS}} \leq \max_{p_X} I(\text{rk}(X); \text{rk}(Y)) \leq \log_2 \min\{M, N, T\}$. ■

Proof of Lemma 6: Let $\tilde{U} = \mathbb{F}^M$. Since $\tilde{V} \leq \tilde{U}$, there exists a full rank $M \times M$ matrix

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_0 \\ \mathbf{D}_1 \end{bmatrix}$$

such that $\langle \mathbf{D}^\top \rangle = \tilde{U}$ and $\langle \mathbf{D}_1^\top \rangle = \tilde{V}$. By Lemma 4,

$$\sum_{s' \geq s} P_{\text{rk}(Y)|\langle X^\top \rangle}(s'|\tilde{V}) = \Pr\{\text{rk}(\mathbf{D}_1 H) \geq s\},$$

and

$$\begin{aligned} P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) &= \Pr\{\text{rk}(\mathbf{D}H) = s\} \\ &= \Pr\{\text{rk}(H) = s\}. \end{aligned}$$

We know $\Pr\{\text{rk}(H) \geq s\} \geq \Pr\{\text{rk}(\mathbf{D}_1 H) \geq s\}$. So

$$\sum_{s' \geq s} P_{\text{rk}(Y)|\langle X^\top \rangle}(s'|\tilde{U}) \geq \sum_{s' \geq s} P_{\text{rk}(Y)|\langle X^\top \rangle}(s'|\tilde{V}).$$

Moreover, for s such that $r < s \leq \text{rk}^*(H)$,

$$\sum_{s': s' \geq s} P_{\text{rk}(Y)|\langle X^\top \rangle}(s'|\tilde{V}) = 0.$$

Thus,

$$\begin{aligned} & \sum_s s(P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) - P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V})) \\ &= \sum_k \sum_{s: s \geq k} (P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) - P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V})) \\ &\geq \sum_{k: \text{rk}^*(H) \geq k > r} \sum_{s: s \geq k} P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) \\ &\geq \sum_{k: \text{rk}^*(H) \geq k > r} \Pr\{\text{rk}(H) = \text{rk}^*(H)\} \\ &= (\text{rk}^*(H) - r) \Pr\{\text{rk}(H) = \text{rk}^*(H)\}. \end{aligned} \tag{28}$$

By definition,

$$\begin{aligned} & \frac{g(\tilde{U}) - g(\tilde{V})}{\log_2 q} \\ &= \sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) \left((T - M)s + \log_q \frac{\zeta_s^T}{\zeta_M^T} \right) \\ & \quad - \sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V}) \left((T - r)s + \log_q \frac{\zeta_s^T}{\zeta_r^T} \right) \\ &= (T - M) \sum_s s(P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) - P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V})) \\ & \quad - (M - r) \sum_s s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V}) \end{aligned}$$

$$\begin{aligned}
& + \sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) \log_q \frac{\zeta_s^T}{\zeta_s^M} \\
& - \sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V}) \log_q \frac{\zeta_s^T}{\zeta_s^r} \\
& > (T - M)(\text{rk}^*(H) - r) \Pr\{\text{rk}(H) = \text{rk}^*(H)\} \\
& - r(M - r) + \log_q \zeta_s^r,
\end{aligned}$$

where the last inequality follows from (28). Therefore

$$\begin{aligned}
(M - r) \sum_s s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V}) & \leq r(M - r), \\
\sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) \log_q \frac{\zeta_s^T}{\zeta_s^M} & \geq 0,
\end{aligned}$$

and

$$\sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V}) \log_q \frac{\zeta_s^T}{\zeta_s^r} < \sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{V}) \log_q \frac{1}{\zeta_s^r} \leq \log_q \frac{1}{\zeta_s^r}.$$

■

Proof of Theorem 13:

We treat $Q_r(\mathbf{X})$ and $R(r)$ as the variables to maximize $I(\langle X \rangle; \langle Y \rangle)$. By the KKT conditions, a set of necessary and sufficient conditions such that an α -type input with variables $Q_r(\mathbf{X})$ and $R(r)$ to achieve $C_{\text{SS}}(H, T)$ is that

$$\begin{aligned}
\frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial Q_r(\tilde{U})} + R(r)g(\tilde{U}) & = \lambda_r \\
\forall r, \tilde{U} \in \text{Gr}(r, \mathbb{F}^M) : Q_r(\tilde{U}) & > 0,
\end{aligned} \tag{29a}$$

$$\begin{aligned}
\frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial Q_r(\tilde{U})} + R(r)g(\tilde{U}) & \leq \lambda_r \\
\forall r, \tilde{U} \in \text{Gr}(r, \mathbb{F}^M) : Q_r(\tilde{U}) & = 0,
\end{aligned} \tag{29b}$$

$$\begin{aligned}
\frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial R(r)} + \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} Q_r(\tilde{U})g(\tilde{U}) & = \bar{\lambda} \\
\forall r : R(r) & > 0,
\end{aligned} \tag{29c}$$

$$\begin{aligned}
\frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial R(r)} + \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} Q_r(\tilde{U})g(\tilde{U}) & \leq \bar{\lambda} \\
\forall r : R(r) & = 0,
\end{aligned} \tag{29d}$$

where the partial derivatives are

$$\begin{aligned}
& \frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial Q_r(\tilde{U})} \\
& = R(r) \sum_s P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\tilde{U}) \log_2 \frac{P_{\text{rk}(Y)|\text{rk}(X)}(s|r)}{P_{\text{rk}(Y)}(s)} - \log_2 e,
\end{aligned}$$

and

$$\begin{aligned} & \frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial R(r)} \\ &= \sum_s P_{\text{rk}(Y)|\text{rk}(X)}(s|r) \log_2 \frac{P_{\text{rk}(Y)|\text{rk}(X)}(s|r)}{P_{\text{rk}(Y)}(s)} - \log_2 e. \end{aligned}$$

We can check that

$$C_{\text{SS}}(H, T) = \bar{\lambda} + \log_2 e,$$

and

$$\bar{\lambda} = \sum_r \lambda_r + (M-1) \log_2 e.$$

To prove the theorem, we only need to check that the α -type input with $R(M) = 1$ satisfies (29). Conditions (29a) and (29b) with $r < M$ are satisfied by $\lambda_r = -\log_2 e$ because $R(r) = 0$. Since $Q_M(\mathbb{F}^M) = 1$, we check condition (29a) with $r = M$. Since $P_{\text{rk}(Y)|\text{rk}(X)}(s|M) = P_{\text{rk}(Y)}(s)$,

$$\left. \frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial Q_M(\mathbb{F}^M)} \right|_{R(M)=1} = -\log_2 e.$$

So, (29a) with $r = M$ is satisfied by $\lambda_M = g(\mathbb{F}^M) - \log_2 e$. This completes the verification of (29a) and (29b).

The above analysis also tells that $\bar{\lambda} = \lambda_M$. Now we check (29c) and (29d) with $\bar{\lambda} = g(\mathbb{F}^M) - \log_2 e$. Since $R(M) = 1$, condition (29c) should be satisfied with $r = M$. This is true since

$$\left. \frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial R(M)} \right|_{R(M)=1} + g(\mathbb{F}^M) = -\log_2 e + g(\mathbb{F}^M).$$

Next, we check condition (29d) for $r < M$. We know

$$\begin{aligned} & \left. \frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial R(r)} \right|_{R(M)=1} \\ &= \underbrace{\sum_s P_{\text{rk}(Y)|\text{rk}(X)}(s|r) \log_2 \frac{P_{\text{rk}(Y)|\text{rk}(X)}(s|r)}{P_{\text{rk}(Y)|\text{rk}(X)}(s|M)}}_{(A)} - \log_2 e. \end{aligned}$$

Since

$$\begin{aligned} P_{\text{rk}(Y)|\text{rk}(X)}(s|M) &= P_{\text{rk}(Y)|\langle X^\top \rangle}(s|\mathbb{F}^M) \\ &= \Pr\{\text{rk}(\mathbf{D}H) = s\} \\ &= \Pr\{\text{rk}(H) = s\}, \end{aligned}$$

we have

$$\begin{aligned} (A) &\leq \sum_s P_{\text{rk}(Y)|\text{rk}(X)}(s|r) \log_2 \frac{1}{P_{\text{rk}(Y)|\text{rk}(X)}(s|M)} \\ &= \sum_s P_{\text{rk}(Y)|\text{rk}(X)}(s|r) \log_2 \frac{1}{p_{\text{rk}(H)}(s)} \\ &\leq -\log_2 \min_{0 \leq s < M} p_{\text{rk}(H)}(s). \end{aligned}$$

That is

$$\left. \frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial R(r)} \right|_{R(M)=1} \leq -\log_2 \min_{0 \leq s \leq M} p_{\text{rk}(H)}(s) - \log_2 e.$$

Fix T_1 such that $\Theta(T_1, r, H) \geq -\log_2 \min_{0 \leq s \leq M} p_{\text{rk}(H)}(s)$ for all $r < M$. This is possible because $\Theta(T, r, H)$ is linearly increasing with T and $-\log_2 \min_{0 \leq s \leq M} p_{\text{rk}(H)}(s)$ does not change with T . By Lemma 6, $g(\mathbb{F}^M) \geq g(\tilde{U}) - \log_2 \min_{0 \leq s \leq M} p_{\text{rk}(H)}(s)$ for all $\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)$. Thus

$$\begin{aligned} \bar{\lambda} &= g(\mathbb{F}^M) - \log_2 e \\ &\geq \max_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} g(\tilde{U}) - \log_2 \min_{0 \leq s \leq M} p_{\text{rk}(H)}(s) - \log_2 e \\ &\geq \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} Q_r(\tilde{U}) g(\tilde{U}) + \left. \frac{\partial I(\text{rk}(X); \text{rk}(Y))}{\partial R(r)} \right|_{R(M)=1}. \end{aligned}$$

Hence, condition (29d) with $r < M$ is satisfied. ■

VIII. MORE ABOUT UNIFORM LOCs

A. Alternative Definition

We have the following alternative definition of uniform LOCs, where the equivalence follows from the symmetric property in Theorem 1 (see proof in Lemma 7).

Definition 6 (Alternative definition of uniform LOCs): A LOC (H, T) is *uniform* if there exists a function $\mu_{\text{rk}} : \mathbb{Z}^+ \times \mathbb{Z}^+ \rightarrow [0, 1]$ such that

$$\Pr\{\mathbf{Y} = \mathbf{X}H\} = \begin{cases} \mu_{\text{rk}}(\text{rk}(\mathbf{X}), \text{rk}(\mathbf{Y})) & \langle \mathbf{Y} \rangle \subseteq \langle \mathbf{X} \rangle \\ 0 & o.w. \end{cases}$$

Lemma 7: Consider a uniform LOC (H, T) and the function μ defined in Definition 3. For $V \leq U \leq \mathbb{F}^T$ and $V' \leq U' \leq \mathbb{F}^T$ with $\dim(V') = \dim(V)$ and $\dim(U') = \dim(U)$, $\mu(U', V') = \mu(U, V)$.

Proof: Let $\dim(V) = s$ and $\dim(U) = r$. Find matrices \mathbf{B} and \mathbf{B}' with $T \times s$ such that $\langle \mathbf{B} \rangle = V$ and $\langle \mathbf{B}' \rangle = V'$. Furthermore, there exists $\mathbf{X} = [\mathbf{B} \quad \mathbf{C}]$ and $\mathbf{X}' = [\mathbf{B}' \quad \mathbf{C}']$ such that $\langle \mathbf{X} \rangle = U$ and $\langle \mathbf{X}' \rangle = U'$. There exists $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$ such that $\Phi \mathbf{B} = \mathbf{B}'$ and $\Phi \mathbf{X} = \mathbf{X}'$ (ref. the proof of Lemma 16 in Appendix B). Thus,

$$\begin{aligned} p_{\langle Y \rangle | X}(V | \mathbf{X}) &= \sum_{\mathbf{Y} : \langle \mathbf{Y} \rangle = V} P_{Y|X}(\mathbf{Y} | \mathbf{X}) \\ &= \sum_{\mathbf{Y} : \langle \mathbf{Y} \rangle = V} P_{Y|X}(\Phi \mathbf{Y} | \Phi \mathbf{X}) \\ &= \sum_{\mathbf{Y}' : \langle \mathbf{Y}' \rangle = \Phi V} P_{Y|X}(\mathbf{Y}' | \mathbf{X}') \\ &= p_{\langle Y \rangle | X}(V' | \mathbf{X}'), \end{aligned} \tag{30}$$

where (30) follows from Theorem 1. On the other hand,

$$\begin{aligned}
 p_{\langle Y \rangle | X}(V | \mathbf{X}) &= \sum_{\mathbf{Y}: \langle \mathbf{Y} \rangle = V} P_{Y|X}(\mathbf{Y} | \mathbf{X}) \\
 &= \sum_{\mathbf{Y}: \langle \mathbf{Y} \rangle = V} \mu(U, V) \\
 &= \chi_s^N \mu(U, V),
 \end{aligned}$$

and similarly

$$p_{\langle Y \rangle | X}(V | \mathbf{X}) = \chi_s^N \mu(U', V').$$

Therefore, $\mu(U', V') = \mu(U, V)$. ■

B. Simplification for Uniform LOCs

For uniform LOCs, the discussion in Section VI and VII can be further simplified. For a uniform LOC, let $\mathbf{X} \in \mathbb{F}^{T \times M}$ with $\langle \mathbf{X}^\top \rangle = \tilde{U}$. By Lemma 4

$$\begin{aligned}
 P_{\text{rk}(Y) | \langle X^\top \rangle}(s | \tilde{U}) &= P_{\text{rk}(Y) | X}(s | \mathbf{X}) \\
 &= \sum_{\mathbf{Y}: \langle \mathbf{Y} \rangle \subset \langle \mathbf{X} \rangle, \text{rk}(Y)=s} P_{Y|X}(\mathbf{Y} | \mathbf{X}) \\
 &= \sum_{\mathbf{Y}: \langle \mathbf{Y} \rangle \subset \langle \mathbf{X} \rangle, \text{rk}(Y)=s} \mu_{\text{rk}}(\dim(\tilde{U}), s) \\
 &= \binom{\dim(\tilde{U})}{s}_q \mu_{\text{rk}}(\dim(\tilde{U}), s) \\
 &\triangleq \tilde{\mu}(\dim(\tilde{U}), s).
 \end{aligned} \tag{31}$$

Using (31), we can rewrite (15) as

$$\begin{aligned}
 P_{\text{rk}(Y) | \text{rk}(X)}(s | r) &= \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} P_{\text{rk}(Y) | \langle X^\top \rangle}(s | \tilde{U}) Q_r(\tilde{U}) \\
 &= \sum_{\tilde{U} \in \text{Gr}(r, \mathbb{F}^M)} \tilde{\mu}(r, s) Q_r(\tilde{U}) \\
 &= \tilde{\mu}(r, s).
 \end{aligned}$$

So $I(\text{rk}(X); \text{rk}(Y))$ in (16) is only a function of $R(r)$. Let

$$g(r) = \sum_s \tilde{\mu}(r, s) \log_2 \frac{\chi_s^T}{\chi_s^r}.$$

Using (31), we have $g(\tilde{U})$ defined in (17) satisfying $g(\tilde{U}) = g(\text{rk}(\tilde{U}))$.

$$J(\text{rk}(X); \text{rk}(Y)) = \sum_r R(r) g(r).$$

Thus, both $I(\text{rk}(X); \text{rk}(Y))$ and $J(\text{rk}(X); \text{rk}(Y))$ are only a function of $R(r)$. So to maximize $I(\langle X \rangle; \langle Y \rangle)$ of a uniform LOC, we only need to consider the rank distribution of an α -type input and the distribution Q_r can be arbitrarily chosen.

Other results in Section VI and VII can be accordingly simplified and we do not repeat the procedure here.

C. Proof of Theorem 6

Proof of i). Fix $\mathbf{X} \in \mathbb{F}^{T \times M}$ with $\dim(\mathbf{X}) = M$. The existence of such \mathbf{X} follows from $T \geq M$. Thus, for any $\mathbf{Y} \in \mathbb{F}^{T \times N}$, we have unique \mathbf{H} such that $\mathbf{Y} = \mathbf{X}\mathbf{H}$. So

$$\begin{aligned} p_H(\mathbf{H}) &= \Pr\{\mathbf{Y} = \mathbf{X}\mathbf{H}\} \\ &= \mu_{\text{rk}}(\text{rk}(\mathbf{X}), \text{rk}(\mathbf{Y})) \\ &= \mu_{\text{rk}}(M, \text{rk}(\mathbf{H})), \end{aligned}$$

where μ_{rk} is given by the alternative definition of uniform LOCs. Therefore H is uniform.

Proof of ii). Let $\mathbf{X} \in \mathbb{F}^{T \times M}$ and $\mathbf{Y} \in \mathbb{F}^{T \times N}$ with $\text{rk}(\mathbf{X}) = r$, $\text{rk}(\mathbf{Y}) = s$ and $\langle \mathbf{Y} \rangle \subset \langle \mathbf{X} \rangle$. Fix a full rank decomposition $\mathbf{X} = \mathbf{B}\mathbf{D}$ and write $\mathbf{Y} = \mathbf{B}\mathbf{E}$. By Prop. 1, we have

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \Pr\{\mathbf{D}\mathbf{H} = \mathbf{E}\}.$$

Let $A = \{\mathbf{H} \in \mathbb{F}^{M \times N} : \mathbf{D}\mathbf{H} = \mathbf{E}\}$. For any $\Psi \subset \mathbb{F}^{M \times N}$, let

$$N_\Psi(k) = |\{\mathbf{H} \in \Psi : \text{rk}(\mathbf{H}) = k\}|.$$

Since H is uniform, we have

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \sum_k N_A(k) \frac{p_{\text{rk}(H)}(k)}{\chi_k^{M,N}}. \quad (32)$$

To finish the proof, we need to determine $N_A(k)$.

Without loss of generality, we assume that the first r columns of \mathbf{D} are linearly independent. Write $\mathbf{D} = [\mathbf{D}_1 \ \mathbf{D}_2]$. We have $\mathbf{D}\mathbf{H}_0 = \mathbf{E}$, where

$$\mathbf{H}_0 = \begin{bmatrix} \mathbf{D}_1^{-1} \mathbf{E} \\ \mathbf{0} \end{bmatrix}.$$

We know $\text{rk}(\mathbf{H}_0) = \text{rk}(\mathbf{E}) = s$. Find $\Phi \in \text{Fr}(\mathbb{F}^{N \times N})$ such that

$$\mathbf{H}_0 \Phi = \begin{bmatrix} \mathbf{H}_{00} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{H}}_0 & \mathbf{0} \end{bmatrix},$$

where \mathbf{H}_{00} is an $r \times s$ full rank matrix and $\tilde{\mathbf{H}}_0$ is the first r columns of $\mathbf{H}_0 \Phi$. Since $\text{rk}(\mathbf{D}) = r$, the null space of \mathbf{D} , defined as

$$\text{Null}(\mathbf{D}) = \{\mathbf{h} \in \mathbb{F}^{M \times 1} : \mathbf{D}\mathbf{h} = \mathbf{0}\},$$

has dimension $M - r$. Let $\mathbf{g}_i, i = 1, 2, \dots, (M - r)$, be a basis of the null space of \mathbf{X} , and let \mathbf{G} be the matrix formed by juxtaposing the vectors in the basis. Thus,

$$\mathbf{A} = \mathbf{H}_0 + \mathbf{G}\mathbb{F}^{(M-r) \times N}.$$

Since $\text{rk}(\mathbf{H}) = \text{rk}(\mathbf{H}\Phi)$ for all $\mathbf{H} \in \mathbb{F}^{M \times N}$,

$$N_A(k) = N_{A\Phi}(k),$$

where

$$A\Phi = \mathbf{H}_0\Phi + \mathbf{G}\mathbb{F}^{(M-r) \times N}.$$

Now we study the rank of $\mathbf{H} = \mathbf{H}_0\Phi + \mathbf{G}\mathbf{F}$ for $\mathbf{F} \in \mathbb{F}^{(M-r) \times N}$. We know

$$\mathbf{H} = \begin{bmatrix} \tilde{\mathbf{H}}_0 + \mathbf{G}\mathbf{F}_1 & \mathbf{G}\mathbf{F}_2 \end{bmatrix},$$

where $\mathbf{F} = [\mathbf{F}_1 \ \mathbf{F}_2]$, $\mathbf{F}_1 \in \mathbb{F}^{(M-r) \times s}$ and $\mathbf{F}_2 \in \mathbb{F}^{(M-r) \times (N-s)}$. We show $\langle \mathbf{H}_0 \rangle \cap \langle \mathbf{G} \rangle = \mathbf{0}$. If there exists nonzero $\mathbf{h} \in \langle \mathbf{H}_0 \rangle \cap \langle \mathbf{G} \rangle$, by the structure of \mathbf{H}_0 , we have

$$\mathbf{h} = \begin{bmatrix} \tilde{\mathbf{h}} \\ \mathbf{0} \end{bmatrix}.$$

Since \mathbf{h} is in the null space of \mathbf{D} , $\mathbf{0} = \mathbf{D}\mathbf{h} = \mathbf{D}_1\tilde{\mathbf{h}}$. Because \mathbf{D}_1 is full rank, we have $\tilde{\mathbf{h}} = \mathbf{0}$. This is a contradiction to $\mathbf{h} \neq \mathbf{0}$. Therefore, $\langle \mathbf{H}_0 \rangle \cap \langle \mathbf{G} \rangle = \mathbf{0}$. Hence, $\langle \tilde{\mathbf{H}}_0 \rangle \cap \langle \mathbf{G} \rangle = \mathbf{0}$, which implies $\text{rk}(\tilde{\mathbf{H}}_0 + \mathbf{G}\mathbf{F}_1) = s$. Thus,

$$\begin{aligned} \text{rk}(\mathbf{H}) &= \text{rk}(\tilde{\mathbf{H}}_0 + \mathbf{G}\mathbf{F}_1) + \text{rk}(\mathbf{G}\mathbf{F}_2) \\ &= s + \text{rk}(\mathbf{F}_2). \end{aligned}$$

This gives that

$$\begin{aligned} N_{A\Phi}(k) &= |\mathbb{F}^{(M-r) \times s}| N_{\mathbb{F}^{(M-r) \times (N-s)}}(k - s) \\ &= \chi_{k-s}^{(M-r), (N-s)} q^{(M-r)s}. \end{aligned}$$

Taking $N_A(k) = \chi_{k-s}^{(M-r), (N-s)}$ into (32), we have

$$P_{Y|X}(\mathbf{Y}|\mathbf{X}) = \sum_{k \geq s} p_{\text{rk}(H)}(k) \frac{\chi_{k-s}^{(M-r), (N-s)} q^{(M-r)s}}{\chi_s^{M, N}}.$$

Let

$$\mu(r, s) = \sum_{k \geq s} p_{\text{rk}(H)}(k) \frac{\chi_{k-s}^{(M-r), (N-s)} q^{(M-r)s}}{\chi_s^{M, N}}.$$

Therefore, (H, T) is uniform. The proof is complete.

IX. CONCLUDING REMARKS

Linear operator channels with arbitrarily distributed transfer matrices are studied. One important guideline we obtained here is that using constant-dimensional subspace coding suffices if we want to use subspace coding for LOCs. We give the method to find the subspaces for constant-dimensional subspace coding. When the packet length is short, encoding/decoding techniques for LOCs still need further investigations.

APPENDIX A

COUNTING

Parts of the counting problems here can be found in various sources, e.g., [14] and reference therein. Here we give the self-contained proofs.

The *projective space* $\text{Pj}(\mathbb{F}^t)$ is the collection of all subspaces of \mathbb{F}^t . Let $\text{Pj}(m, \mathbb{F}^t)$ be the subset of $\text{Pj}(\mathbb{F}^t)$ that contains all the subspaces with dimension less than or equal to m . Let $\text{Fr}(\mathbb{F}^{m \times r})$ be the set of full rank matrices in $\mathbb{F}^{m \times r}$. Define

$$\chi_r^m = \begin{cases} (q^m - 1)(q^m - q) \cdots (q^m - q^{r-1}) & r > 0 \\ 1 & r = 0 \end{cases} \quad (33)$$

for $r \leq m$.

Lemma 8: When $0 \leq r \leq m$, $|\text{Fr}(\mathbb{F}^{m \times r})| = \chi_r^m$

Proof: The lemma is trivial for $r = 0$, so we consider $r > 0$. We can count the number of full rank matrices in $\mathbb{F}^{m \times r}$ by the columns. For the first column, we can choose all vectors in \mathbb{F}^m except the zero vector. Thus we have $q^m - 1$ choices. Fixed the first column, say v_1 , we want to choose the second column v_2 in \mathbb{F}^m but is linear independent with v_1 . Hence, we have $q^m - q$ choices of v_2 . Repeat this process, we obtain that the number of full rank $m \times r$ matrices is $(q^m - 1)(q^m - q) \cdots (q^m - q^{r-1}) = \chi_r^m$. ■

Define

$$\zeta_r^m = \chi_r^m q^{-mr}. \quad (34)$$

Since the number of $m \times r$ matrices is q^{mr} , ζ_r^m can be regarded as the probability that a randomly chosen $m \times r$ matrix is full rank.

Lemma 9: Let G be an $s \times m$ random matrix with uniformly independent components over \mathbb{F} . Then for $r \leq m$,

$$p_{\text{rk}(GH) | \text{rk}(H)}(s | r) = \zeta_r^s,$$

where H is any $m \times n$ random matrix.

Proof: Fix an $m \times n$ matrix \mathbf{H} with $\text{rk}(\mathbf{H}) = r$. Let $F = G\mathbf{H}$ and let g_i and f_i be the i th row of G and F , respectively. Since g_i contains uniformly independent components, $\Pr\{g_i = \mathbf{g}\} = q^{-m}$. For \mathbf{f} with $\mathbf{f}^\top \in \langle \mathbf{H}^\top \rangle$,

$$\Pr\{g_i \mathbf{H} = \mathbf{f}\} = q^{-m} |\text{Ker}(\mathbf{H})| = q^{-r},$$

where $\text{Ker}(\mathbf{H}) = \{\mathbf{g} : \mathbf{g}\mathbf{H} = \mathbf{0}\}$ and $|\text{Ker}(\mathbf{H})| = q^{m - \text{rk}(\mathbf{H})}$. So for \mathbf{F} with $\langle \mathbf{F}^\top \rangle \leq \langle \mathbf{H}^\top \rangle$,

$$\begin{aligned} p_{GH|H}(\mathbf{F} | \mathbf{H}) &= \Pr\{g_i \mathbf{H} = \mathbf{f}_i, i = 1, \dots, s\} \\ &= \prod_{i=1}^s \Pr\{g_i \mathbf{H} = \mathbf{f}_i\} \\ &= q^{-sr}. \end{aligned} \quad (35)$$

Thus,

$$\begin{aligned}
p_{\text{rk}(GH)|H}(s|\mathbf{H}) &= q^{-mr} |\{\mathbf{F} : \langle \mathbf{F}^\top \rangle \leq \langle \mathbf{H}^\top \rangle, \text{rk}(\mathbf{F}) = s\}| \\
&= q^{-mr} \chi_s^r \\
&= \zeta_s^r,
\end{aligned}$$

where $|\{\mathbf{F} : \langle \mathbf{F}^\top \rangle \leq \langle \mathbf{H}^\top \rangle, \text{rk}(\mathbf{F}) = s\}| = \chi_s^r$ follows from Lemma 8 in Appendix A. Last, since $\text{rk}(H) \rightarrow H \rightarrow \text{rk}(GH)$ forms a Markov chain,

$$\begin{aligned}
p_{\text{rk}(GH)|\text{rk}(H)}(s|r) &= \sum_{\mathbf{H}:\text{rk}(\mathbf{H})=r} p_{\text{rk}(GH)|H}(s|\mathbf{H}) p_{H|\text{rk}(H)}(\mathbf{H}|r) \\
&= \zeta_s^r \sum_{\mathbf{H}:\text{rk}(\mathbf{H})=r} p_{H|\text{rk}(H)}(\mathbf{H}|r) \\
&= \zeta_s^r.
\end{aligned}$$

The proof is complete. ■

The *Grassmannian* $\text{Gr}(r, \mathbb{F}^t)$ is the set of all r -dimensional subspaces of \mathbb{F}^t . Thus $\text{Pj}(m, \mathbb{F}^t) = \bigcup_{r \leq m} \text{Gr}(r, \mathbb{F}^t)$. The *Gaussian binomials* are defined as

$$\binom{m}{r}_q = \frac{\chi_r^m}{\chi_r^r}.$$

Lemma 10: The number of r -dimensional subspace in \mathbb{F}^m is given by the *Gaussian binomials*.

Proof: Define an equivalent relation on $\mathcal{M}(\mathbb{F}^{m \times r})$ by $\mathbf{X} \sim \mathbf{X}'$ if $\langle \mathbf{X} \rangle = \langle \mathbf{X}' \rangle$. The equivalent class $[\mathbf{X}]$ is the set of all matrices that equivalent to \mathbf{X} . We have $[\mathbf{X}] = \{\mathbf{X}\Phi : \Phi \in \mathcal{M}(\mathbb{F}^{r \times r})\}$. Thus $|[\mathbf{X}]| = |\mathcal{M}(\mathbb{F}^{r \times r})| = \chi_r^r$. Since $\text{Gr}(r, \mathbb{F}^T) = \mathcal{M}(\mathbb{F}^{m \times r}) / \sim$, the quotient set of $\mathcal{M}(\mathbb{F}^{m \times r})$ by \sim , we have $|\text{Gr}(r, \mathbb{F}^T)| = |\mathcal{M}(\mathbb{F}^{m \times r})| / |[\mathbf{X}]| = \chi_r^m / \chi_r^r$. ■

Let

$$\chi_r^{m,n} = \frac{\chi_r^m \chi_r^n}{\chi_r^r}, \tag{36}$$

which is the number of $m \times n$ matrices with rank r .

Lemma 11: For $m \geq r'$ and $r \geq r'$, define a set $S = \{\mathbf{X} \in \mathbb{F}^{m \times r} : \text{rk}(\mathbf{X}) = r'\}$. Then

$$|S| = \frac{\chi_{r'}^m \chi_{r'}^r}{\chi_{r'}^{r'}} = \chi_{r'}^{m,r}. \tag{37}$$

Furthermore,

$$\sum_{r'} \chi_{r'}^{m,r} = q^{mr}. \tag{38}$$

Proof: The column vectors of $\mathbf{X} \in S$ span an r' -dimensional subspace in an m -dimensional vector space. Let $\{V_1, V_2, \dots, V_n\}$ be the set of r' -dimensional subspace in an m -dimensional vector space, where $n = \binom{m}{r'}_q$. Let $S_{V_i} = \{\mathbf{X} \in \mathbb{F}^{m \times r} : \langle \mathbf{X} \rangle = V_i\}$ and the set $\{S_{V_i}\}$ is a partition of S . We know that $|\{S_{V_i}\}| = \chi_{r'}^r$. Therefore,

$$|S| = n |S_{V_i}| = \binom{m}{r'}_q \chi_{r'}^r = \chi_{r'}^{m,r}. \tag{39}$$

The equality in (38) follows because both sides are the number of $m \times r$ matrices. ■

Lemma 12: Let $V \leq \mathbb{F}^m$ be an s -dimensional subspace. Then, the number of subspace U with $V \leq U$ and $\dim(U) = r$ is

$$\binom{m-s}{r-s}_q = \binom{m}{r}_q \frac{\chi_s^r}{\chi_s^m}. \quad (40)$$

Proof: Let U be a subspace with $V \leq U$ and $\dim(U) = r$. Then we can write $U = V + U'$ where U' is a $\dim(U') = r - s$ and $V \cap U' = \{0\}$. Given U , such U' is unique. The number of U' is the number of $(r - s)$ -dimensional subspace in an $(m - s)$ -dimensional space, i.e., $\binom{m-s}{r-s}_q$. The equality in (40) is the direct result of the definitions. ■

APPENDIX B USEFUL RESULTS

Define

$$\Xi_q(s) = \prod_{i=s}^{\infty} (1 - q^{-i}). \quad (41)$$

Lemma 13: For $r \leq m$, $\zeta_r^m > \Xi_2(1) > 0.2887$.

Proof: We know $\Xi_q(s+1) > \Xi_q(s) > \Xi_{q-1}(s) \geq \Xi_2(1)$, where $\Xi_2(1)$ is a mathematics constant with approximate value 0.28879 [14]. So $\zeta_r^m > \Xi_q(m-r+1) \geq \Xi_2(1)$. ■

Lemma 14: $\binom{m}{r}_q < cq^{(m-r)r}$, where $c \approx 3.4627$.

Proof: By definition, $\binom{m}{r}_q = q^{(m-r)r} \frac{\zeta_r^m}{\zeta_r^m} < q^{(m-r)r} \frac{1}{\Xi_q(1)} = cq^{(m-r)r}$, where $c = 1/\Xi_2(1) \approx 1/0.28879$. ■

Lemma 15: For $V \leq U \leq \mathbb{F}^T$ and $V' \leq U' \leq \mathbb{F}^T$ with $\dim(U) = \dim(U')$ and $\dim(V) = \dim(V')$, there exists $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$ such that $\Phi U = U'$ and $\Phi V = V'$.

Proof: Find a basis $\{\mathbf{b}_i : i = 1, \dots, T\}$ of \mathbb{F}^T such that $\{\mathbf{b}_i : i = 1, \dots, r\}$ is a basis of U and $\{\mathbf{b}_i : i = 1, \dots, s\}$ is a basis of V . We can do this by first finding a basis of V , extending the basis to a basis of U and further extending to a basis of \mathbb{F}^T . Similarly, find a basis $\{\mathbf{b}'_i : i = 1, \dots, T\}$ of \mathbb{F}^T such that $\{\mathbf{b}'_i : i = 1, \dots, r\}$ is a basis of U' and $\{\mathbf{b}'_i : i = 1, \dots, s\}$ is a basis of V' . Consider the linear system of equations

$$\Phi \mathbf{b}_i = \mathbf{b}'_i, \quad i = 1, \dots, T.$$

We know that there exists a unique $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$ satisfying this linear system and $\Phi V = V'$ and $\Phi U = U'$. ■

Lemma 16: For $\mathbf{X}, \mathbf{X}' \in \mathbb{F}^{T \times M}$, $\langle \mathbf{X}^\top \rangle = \langle \mathbf{X}'^\top \rangle$ if and only if there exists $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$ such that $\mathbf{X}' = \Phi \mathbf{X}$.

Proof: Let $r = \text{rk}(\mathbf{X})$. First, show a) \Rightarrow c). Fix one full-rank decomposition $\mathbf{X} = \mathbf{B}\mathbf{D}$. Since $\langle \mathbf{D}^\top \rangle = \langle \mathbf{X}^\top \rangle = \langle \mathbf{X}'^\top \rangle$, there exists a decomposition $\mathbf{X}' = \mathbf{B}'\mathbf{D}$ using the same procedure we described by first fixing \mathbf{D} . Second, show c) \Rightarrow b). With the decomposition in c), there exists $\Phi \in \text{Fr}(\mathbb{F}^{T \times T})$ such that $\Phi \mathbf{B} = \mathbf{B}'$. Extend \mathbf{B} and \mathbf{B}' to $T \times T$ matrices $[\mathbf{B} \ \mathbf{B}_0]$ and $[\mathbf{B}' \ \mathbf{B}'_0]$. Then, $\Phi = [\mathbf{B}' \ \mathbf{B}'_0][\mathbf{B} \ \mathbf{B}_0]^{-1}$ is one such matrix we want since $\Phi[\mathbf{B} \ \mathbf{B}_0] = [\mathbf{B}' \ \mathbf{B}'_0]$. Last, we have b) \Rightarrow a). ■

Lemma 17: For $U \leq \mathbb{F}^t$ with $\dim(U) = r \leq m$, let

$$A(m, U) = \{\mathbf{X} \in \mathbb{F}^{t \times m} : \langle \mathbf{X} \rangle = U\}.$$

Then,

$$|A(m, U)| = \chi_r^m,$$

and for $\Phi \in \text{Fr}(\mathbb{F}^{t \times t})$

$$A(m, \Phi U) = \Phi A(m, U).$$

Proof: Find a $t \times r$ matrix \mathbf{B} with $\langle \mathbf{B} \rangle = U$. Then, we have

$$A(m, U) = \{\mathbf{B}\mathbf{D} : \mathbf{D} \in \text{Fr}(\mathbb{F}^{r \times m})\} = \mathbf{B} \text{Fr}(\mathbb{F}^{r \times m}).$$

Thus, $|A(m, U)| = |\text{Fr}(\mathbb{F}^{r \times m})| = \chi_r^m$. For $\Phi \in \text{Fr}(\mathbb{F}^{t \times t})$, $\langle \Phi \mathbf{B} \rangle = \Phi U$. So $A(m, \Phi U) = \Phi \mathbf{B} \text{Fr}(\mathbb{F}^{r \times m}) = \Phi A(m, U)$. ■

ACKNOWLEDGEMENT

Shenghao Yang thanks Dr. Kenneth Shum and Prof. Raymond Yeung for helpful discussions.

REFERENCES

- [1] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.
- [2] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Trans. Networking*, vol. 11, no. 5, pp. 782–795, Oct. 2003.
- [3] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3579–3591, Aug. 2008.
- [4] D. Silva, F. R. Kschischang, and R. Koetter, "Communication over finite-field matrix channels," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 1296–1305, Mar. 2010.
- [5] M. Siavoshani, S. Mohajer, C. Fragouli, and S. Diggavi, "On the capacity of noncoherent network coding," *Information Theory, IEEE Transactions on*, vol. 57, no. 2, pp. 1046–1066, feb. 2011.
- [6] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.
- [7] S. Yang, S.-W. Ho, J. Meng, and E. hui Yang, "Optimality of subspace coding for linear operator channels over finite fields," in *Proc. IEEE Information Theory Workshop ITW '10*, Cairo, Egypt, Jan. 2010.
- [8] —, "Linear operator channels over finite fields," 2010. [Online]. Available: <http://arxiv.org/abs/1002.2293v1>
- [9] R. W. Nóbrega, B. F. Uchôa-Filho, and D. Silva, "On the capacity of multiplicative finite-field matrix channels," in *Proc. IEEE ISIT '11*, Saint Petersburg, Russia, August 2011.
- [10] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Oct. 2003.
- [11] P. Maymounkov, N. J. A. Harvey, and D. S. Lun, "Methods for efficient network coding," in *Proc. Allerton Conf. Comm., Control, and Computing*, Sep. 2006.
- [12] R. W. Yeung, *Information Theory and Network Coding*. Springer, 2008.
- [13] S. Yang, J. Meng, and E. hui Yang, "Coding for linear operator channels over finite fields," in *Proc. IEEE Inte. Symp. on Information Theory ISIT '10*, Austin, USA, Jun. 2010.
- [14] C. Cooper, "On the distribution of rank of a random matrix over a finite field," *Random Struct. Algorithms*, vol. 17, no. 3-4, pp. 197–212, 2000.